

VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA  
EKONOMICKÁ FAKULTA

KATEDRA FINANČÍ

**Srovnání skóringových modelů při odhadu pravděpodobnosti úpadku bank  
v USA**

**Comparison of credit scoring models on probability of default estimation  
for US banks**

Student: Bc. Martin Gurný

Vedoucí bakalářské práce: Ing. Tomáš Tichý, Ph.D.

Ostrava 2011

VŠB - Technická univerzita Ostrava  
Ekonomická fakulta  
Katedra financí

## Zadání diplomové práce

Student: **Bc. Martin Gurný**  
Studijní program: N6202 Hospodářská politika a správa  
Studijní obor: 6202T010 Finance  
Specializace: 00 Finance  
Téma: Srovnání skóringových modelů při odhadu pravděpodobnosti úpadku bank v USA  
Comparison of credit scoring models on probability of default estimation for US banks

Zásady pro vypracování:

1. Úvod
  2. Charakteristika modelů odhadu úvěrového rizika
  3. Sestavení skóringových modelů
  4. Aplikace na vybrané banky
  5. Závěr
- Seznam použité literatury  
Seznam zkratk  
Prohlášení o využití výsledků diplomové práce  
Přílohy

Seznam doporučené odborné literatury:

BALTHAZAR, L. *From Basel 1 to Basel 3: The Integration of State-of-the-Art Risk Modeling in Banking Regulation*. 1st ed. New York: Palgrave Macmillan, 2006. 294 s. ISBN 978-1-4039-4888-5.  
HOSMER, D. W.; LEMESHOW, S. *Applied Logistic Regression*. 2nd ed. New York: Wiley, 2000. 375 s. ISBN 978-0-471-35632-8.  
RESTI, A.; SIRONI, A. *Risk management and Shareholders' value in banking*. 1st ed. Chichester: Wiley, 2007. 782 s. ISBN 978-0-470-02978-7.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Tomáš Tichý, Ph.D.**

Datum zadání: 26.11.2010

Datum odevzdání: 29.04.2011

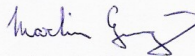
Ing. Iveta Ratmanová, Ph.D.  
vedoucí katedry



prof. Dr. Ing. Dana Dluhošová  
děkanka fakulty

Místopřísežně prohlašuji, že jsem celou práci, včetně všech příloh, vypracoval samostatně.

V Ostravě dne 29. dubna 2011

  
Martin Gurný

Tímto bych chtěl poděkovat Ing. Tomáši Tichému, Ph.D. za odborné vedení, cenné rady, podněty a připomínky poskytnuté při zpracování mé práce a za trpělivost při konzultacích.

Rovněž bych chtěl poděkovat za podporu poskytnutou v rámci řešení projektů Studentské grantové soutěže č. SP2010/102 a SP2011/166, která taktéž vedla ke vzniku této diplomové práce.

<b>1. ÚVOD.....</b>	<b>3</b>
<b>2 CHARAKTERISTIKA MODELŮ ODHADU ÚVĚROVÉHO RIZIKA.....</b>	<b>4</b>
2.1 KREDITNÍ (ÚVĚROVÉ) RIZIKO .....	5
2.1.1 Očekávaná ztráta.....	6
2.1.2 Neočekávaná ztráta .....	7
2.1.3 Pravděpodobnost úpadku (PD) .....	7
2.1.4 Další typy kreditního rizika .....	8
2.2 MODEL PREDIKCE SELHÁNÍ.....	8
2.2.1 Strukturální modely .....	9
2.2.2 Reduced-form modely.....	10
2.2.3 Skóringové modely.....	11
2.3 PŘEHLED DNEŠNÍCH KOMERČNÍCH MODELŮ .....	15
2.3.1 KMV model ( <i>PortfolioManager<sup>TM</sup></i> ).....	16
2.3.2 <i>CreditMetrics<sup>TM</sup></i> .....	17
2.3.3 <i>CreditPortfolioView<sup>TM</sup></i> .....	19
2.3.4 <i>CreditRisk<sup>+</sup></i> .....	19
<b>3 SESTAVENÍ SKÓRINGOVÝCH MODELŮ.....</b>	<b>21</b>
3.1 REGRESNÍ MODEL Y .....	21
3.1.1 Logit model.....	21
3.1.2 Probit model.....	23
3.2 ODHAD PARAMETRŮ A TESTOVÁNÍ STATISTICKÉ VÝZNAMNOSTI LOGIT A PROBIT MODELU .....	24
3.2.1 Odhad parametrů .....	24
3.2.2 Testování významnosti parametrů .....	26
3.2.3 Odhad konfidenčního intervalu .....	28
3.2.4 Stepwise metoda .....	28
3.3 LINEÁRNÍ DISKRIMINAČNÍ ANALÝZA .....	32
3.3.1 Statistická verifikace odhadnutých parametrů a modelu .....	33
3.3.2 Míra úspěšnosti modelu .....	37
3.3.3 Přepočítání skóre na pravděpodobnost úpadku .....	37
3.3.4 Výběr diskriminačních proměnných .....	38
3.4 UKAZATELE POUŽITÉ PŘI SESTAVOVÁNÍ SKÓRINGOVÝCH MODELŮ V APLIKAČNÍ ČÁSTI .....	38
3.4.1 Ukazatele rentability.....	39
3.4.2 Ukazatele výkonnosti .....	40
3.4.3 Ukazatele kvality aktiv.....	40
3.4.4 Ukazatele kapitálové přiměřenosti .....	41
<b>4 APLIKACE NA VYBRANÉ BANKY .....</b>	<b>43</b>
4.1 VSTUPNÍ DATA .....	43
4.1.1 Popis bank obsažených ve vzorku.....	43
4.1.2 Charakteristika finančních ukazatelů .....	44
4.1.3 Korelační analýza ukazatelů.....	45
4.2 LOGIT MODEL.....	45
4.2.1 Odhad logit modelu .....	45
4.2.2 Testování významnosti parametrů .....	47
4.2.3 Grafické znázornění PD jednotlivých bank .....	48
4.3 PROBIT MODEL .....	49
4.3.1 Odhad probit modelu.....	49
4.3.2 Testování významnosti parametrů .....	51
4.3.3 Grafické znázornění PD jednotlivých bank .....	52
4.4 LINEÁRNÍ DISKRIMINAČNÍ ANALÝZA .....	53
4.4.1 Odhad LDA modelu .....	53
4.4.2 Statistická verifikace odhadnutých parametrů a modelu .....	54
4.4.3 Míra úspěšnosti modelu .....	56
4.4.4 Grafické znázornění PD jednotlivých bank .....	56
4.5 SROVNÁNÍ ODHADNUTÝCH MODELŮ .....	57
4.6 APLIKACE ODHADNUTÝCH MODELŮ NA KONTROLNÍM VZORKU BANK .....	59
<b>5 ZÁVĚR.....</b>	<b>63</b>

<b>SEZNAM POUŽITÉ LITERATURY .....</b>	<b>65</b>
<b>SEZNAM ZKRATEK A SYMBOLŮ.....</b>	<b>67</b>
<b>SEZNAM OBRÁZKŮ, TABULEK A GRAFŮ .....</b>	<b>71</b>
<b>PROHLÁŠENÍ O VYUŽITÍ VÝSLEDKŮ DIPLOMOVÉ PRÁCE.....</b>	<b>72</b>
<b>SEZNAM PŘÍLOH.....</b>	<b>73</b>

# 1. Úvod

Klíčovým prvkem moderních finančních trhů je bezesporu určování pravděpodobnosti úpadku finančních subjektů. Žádná z významných oblastí řízení kreditního rizika se neobejde bez správného určení pravděpodobnosti úpadku potenciálního dlužníka. Ať už se jedná o posouzení bonity při žádostech o půjčky, určování ratingu finančního subjektu nebo oceňování kreditních derivátů, nesprávné určení pravděpodobnosti úpadku může mít vážné důsledky na finanční zdraví věřitelů. Zejména špatné oceňování kreditních derivátů hrálo zásadní roli při vzniku nedávné globální ekonomické krize. Podcenění a špatné ocenění rizika mělo za následek kolaps finančního systému, který se právě prostřednictvím kreditních derivátů rozšířil na většinu finančních trhů. Z těchto důvodů je tato práce věnována odhadu pravděpodobnosti úpadku jakožto klíčového parametru v oblasti řízení kreditních rizik.

Cílem diplomové práce je pomocí představené metodologie odhadnout, verifikovat a srovnat modely pro určování pravděpodobnosti úpadku amerických komerčních bank a tyto následně aplikovat na kontrolní vzorek s cílem určit nejvhodnější model.

Teoretická část této práce je rozdělena do dvou kapitol. Ve druhé kapitole jsou nejprve obecně charakterizovány kreditní riziko a základní typy modelů pro predikci selhání (strukturální modely, reduced-form modely a skóringové modely), na konci kapitoly je potom uveden výčet a stručný popis dnešních komerčních modelů pro měření kreditního rizika portfolií.

Třetí kapitola je zaměřena na detailní charakteristiku tří typů skóringových modelů (logit model, probit model a model sestavený na základě lineární diskriminační analýzy) včetně popisu testování statistické významnosti jednotlivých odhadnutých parametrů a modelu jako celku. Součástí této kapitoly je rovněž charakteristika ukazatelů použitých při sestavování modelů v aplikační části.

Čtvrtá kapitola je aplikační a představuje stěžejní část této práce. Její náplní je aplikace představených teoreticko-metodologických východisek na vzorek bank při sestavování jednotlivých typů skóringových modelů. Součástí kapitoly je také charakteristika vstupních dat, srovnání jednotlivých odhadnutých modelů a jejich následná aplikace na kontrolní vzorek bank s cílem určit nejlepší model pro predikci úpadku.

## 2 Charakteristika modelů odhadu úvěrového rizika

Existují různá finanční rizika, kterým musí at' už výrobní podniky či třeba finanční instituce při své existenci čelit. Obecně můžeme riziko označit za měřitelnou možnost, že situace v budoucnu bude jiná než plánovaná. Finanční rizika jsou rizika spojená s finančními aktivitami. Cílem specialistů řízení rizik je udržet riziko v předem stanovené úrovni pomocí vhodných metodik a nástrojů a při této míře rizika se snažit maximalizovat zisk.

Finanční rizika členíme dle toho, jaký mají charakter, na riziko kreditní, tržní, operační a riziko likvidity.

**Kreditní (úvěrové) riziko** je spojeno s nebezpečím, že dlužník nebude schopen dostat svým závazkům a tedy splácet dluhy dle smlouveného splátkového režimu. Tato schopnost je úzce svázána s ratingem, který je dlužníkům přiřazen a charakterizuje stupeň důvěry, že dlužník splní své závazky.

**Tržní riziko** vyplývá ze změn tržních cen a jejich dopadu na zisk firmy. Výše tržního rizika závisí na struktuře bilance a citlivosti jednotlivých položek aktiv a pasiv na změny tržních cen. V závislosti na tom, o jaké tržní ceny se jedná, se tržní riziko dále dělí na riziko měnové, úrokové, akciové, opční a komoditní.

**Riziko likvidity** znamená možnost, že firma ztratí schopnost dostát v každém okamžiku svým splatným závazkům. U finančních institucí to znamená především ztrátu schopnosti kdykoliv vyplatit splatné vklady klientů.

Pod **operačním rizikem** se obecně rozumí možnost vzniku ztráty v důsledku provozních nedostatků a chyb (vnitřních procesů, osob, systémů, externích událostí). V úzkém pojetí lze za operační riziko považovat riziko plynoucí z operací firmy. V širokém pojetí spadají do této kategorie všechna rizika, která nelze přiřadit k riziku kreditnímu, tržnímu nebo riziku likvidity.



## **2.1 Kreditní (úvěrové) riziko**

Kreditní riziko je blíže specifikováno v Duffie a Singleton (2003), Bluhm, Overbeck a Wagner (2003) či Resti a Sironi (2007). Toto riziko můžeme definovat jako možnost, že neočekávaná změna v bonitě protistrany může vyvolat odpovídající nečekanou změnu v tržní hodnotě související kreditní expozice. Kreditní expozicí je myšlena ztráta v případě selhání dlužníka. Tato definice v sobě obsahuje tři koncepty.

### **Riziko úpadku a riziko pravděpodobnosti přechodu**

V první řadě kreditní riziko není omezeno pouze na možnost úpadku protistrany. I pouhé zhoršení bonity dlužníka představuje projev kreditního rizika. Kreditní riziko se tedy skládá převážně ze dvou různých rizik: z rizika úpadku (*default risk*) a z rizika pravděpodobnosti přechodu (*migration risk*). První riziko představuje riziko ztráty vyplývající z aktuální platební neschopnosti dlužníka, zatímco riziko pravděpodobnosti přechodu představuje riziko ztráty vyplývající z pouhého zhoršení bonity dlužníka.

Řízení kreditního rizika by tedy nemělo být založeno na jednoduchém binomickém rozdělení možných událostí („*default*“ a „*non-default*“), ale spíše na spojitém rozdělení, kdy úpadek dlužníka představuje pouze extrémní případ vedle ostatních případů, kdy dlužník sice zůstává solventní, ale pravděpodobnost budoucího úpadku se postupně zvyšuje.

### **Riziko jako nečekaná událost**

Pokud máme mluvit o riziku, změna ratingu dlužníka musí být neočekávaná. Pokud banka poskytne úvěr s vědomím, že u protistrany v budoucnu dojde ke zhoršení kvality (rentability, solventnosti, likvidity atd.), tak toto zhoršení bude muset být vhodně vyhodnoceno a zapracováno do rozhodnutí o poskytnutí úvěru a do tvorby cen (volby úrokové míry). Ve skutečnosti je očekávaný vývoj ekonomické a finanční situace dlužníka vždy brán v úvahu, pokud jsou určeny pravděpodobnost úpadku a související úrokové sazby. Skutečné riziko představuje možnost, že tato hodnocení by se později mohla ukázat jako nesprávná, tedy že u protistrany dojde k nepředvídanému zhoršení. V této souvislosti se vlastní riziko vztahuje pouze na takové události, které ač jsou předvídatelné, tak jsou neočekávané.

## Kreditní expozice

Kreditní riziko není omezeno pouze na „klasické“ formy úvěrů poskytovaných bankami (úvěry a cenné papíry zaznamenané v rozvaze), ale podléhají mu také operace, které v rozvaze nenajdeme. Těmi mohou být různé záruky, deriváty obchodované na OTC (*Over the Counter*) trzích či transakce s cennými papíry nebo se zahraničními měnami.

### 2.1.1 Očekávaná ztráta

Banky se před ztrátou potřebují chránit, ochranu si zajišťují ve formě pojištění. Navíc historie ukazuje, že i dobří dlužníci mají potenciál k neplnění svých finančních závazků. Banky si tedy účtují přiměřené rizikové prémie ke každému úvěru a shromažďují tyto prémie na interním bankovním účtu zvaném rezervy na očekávané ztráty. Tímto si vytvářejí kapitálový polštář pro krytí ztrát z nesplacených úvěrů.

Očekávaná ztráta (**EL** – *expected loss*) je střední hodnota rozdělení pravděpodobnosti budoucích ztrát. Výpočet očekávané ztráty z kreditní expozice vyžaduje odhadnutí následujících tří parametrů:

- a) očekávaná hodnota expozice v případě selhání (**EAD** – *exposure at default*), což je náhodná proměnná vyjádřená současnou expozicí plus možnou změnou ve výši půjčky, která může trvat ode dneška do data možného selhání;
- b) pravděpodobnost, že dlužník zbankrotuje (**PD** – *probability of default*);
- c) očekávaná ztrátovost v případě selhání (**LGD** – *loss given default*), která popisuje část ztrát, které banka v případě selhání dlužníka skutečně utrpí; vypočte se jako rozdíl jedné a míry návratnosti (**RR** – *recovery rate*).

Očekávanou ztrátu tedy můžeme vypočítat dle následujícího vztahu, kde mezi sebou tyto tři parametry násobíme.

$$EL = \overline{EAD} \cdot PD \cdot \overline{LGD} \quad (2.1)$$

### 2.1.2 Neočekávaná ztráta

Držení kapitálu jako polštáře proti očekávaným ztrátám nestačí. Ve skutečnosti by banka měla kromě rezerv na očekávané ztráty také šetřit peníze na pokrytí neočekávané ztráty přesahující průměrné ztráty známé z minulosti. Skutečné kreditní riziko (riziko, že se ztráta ukáže větší, než bylo původně odhadováno) je totiž spojeno právě s neočekávanou ztrátou. Neočekávanou ztrátu můžeme definovat jako variabilitu ztráty kolem střední hodnoty očekávané ztráty.

Rozdíl mezi očekávanou a neočekávanou ztrátou je důležitý z pohledu diverzifikovaného portfolia expozic. Očekávaná ztráta plynoucí z takového portfolia je rovna součtu očekávaných ztrát jednotlivých úvěrů v něm obsažených, zatímco volatilita ztráty celkového portfolia je obecně nižší než součet volatilit ztrát z jednotlivých úvěrů. Jinými slovy zatímco očekávaná ztráta nemůže být redukována diverzifikací portfolia, neočekávanou ztrátu lze snížit vhodnou diverzifikací (rozdělením rizika do různých odvětví, zemí atd.). To znamená, že efektivní diverzifikací úvěrového portfolia lze výrazně snížit celkové kreditní riziko.

### 2.1.3 Pravděpodobnost úpadku (PD)

Přiřazování pravděpodobnosti úpadku ke každému dlužníkovi v úvěrovém portfoliu banky není snadné. V zásadě existují dva přístupy, jak pravděpodobnost úpadku určit.

#### Určení pravděpodobnosti úpadku pomocí tržních dat

Nejznámějším zástupcem tohoto typu určování pravděpodobnosti úpadku je koncept očekávané frekvence úpadku (**EDF** – *expected default frequencies*), jež byl vytvořen firmou KMV Corporation.

Další metoda pro určování pravděpodobnosti úpadku pomocí tržních dat je založena na kreditním rozpětí obchodovaných produktů nesoucích kreditní riziko, jako jsou korporátní dluhopisy a kreditní deriváty (např. credit default swaps).

## Určení pravděpodobnosti úpadku pomocí ratingu

V tomto přístupu jsou pravděpodobnosti úpadku spojeny s ratingem. Rating je dlužníkům přiřazen buď externími ratingovými agenturami, jako jsou *Moody's Investors Services* či *Standard & Poor's*, nebo pomocí interní bankovní metodologie pro určování ratingu dlužníků.

### 2.1.4 Další typy kreditního rizika

Úvěrové riziko se kromě rizika úpadku a rizika pravděpodobnosti přechodu skládá z několika dalších typů rizik.

**Riziko rozpětí** (*spread risk*): riziko spojené se změnou postoje investorů k riziku, v důsledku čehož dochází ke změně rizikové přírážky pro danou míru rizika;

**riziko náhrady** (*recovery risk*): riziko spojené se změnou míry návratnosti v případě selhání dlužníka v důsledku nových informací, změny hladiny cen v případě zajištění, změny odhadované délky likvidačního řízení apod.;

**riziko nahrazení** (*pre-settlement or substitution risk*): riziko insolvence protistrany v případě OTC kontraktů s deriváty; v případě této události může být banka nucena najít pro kontrakt novou protistranu (i za potenciálně méně příznivých podmínek);

**riziko země** (*country risk*): riziko spojené s politickými či legislativními událostmi, které ovlivní možnost převodu prostředků.

Tato klasifikace je založena na tzv. *driverech* základního rizika (selhání, zhoršení bonity, zvýšení averze k riziku, snížení míry návratnosti) a na povaze expozice (OTC kontrakty s deriváty, zahraniční dlužníci).

## 2.2 Modely predikce selhání

Existují dva způsoby, jak predikovat selhání dlužníka. První způsob vychází z ekonomických a finančních ukazatelů. Jedná se o tzv. skóringové modely (*credit scoring models*), které studují významnost vybraných ukazatelů pro finanční zdraví subjektu pomocí

rozličných metod, jako diskriminační analýza, regresní modely (lineární, logit, probit) nebo induktivní modely (neuronové sítě, genetické algoritmy).

Druhým způsobem, jak odhadnout pravděpodobnost selhání, je využít tržního ocenění. Základním předpokladem tedy jsou racionálně se chovající účastníci trhu disponující veškerými relevantními informacemi, které využijí při stanovení požadované výnosnosti. Pak je zřejmé, že výnos rizikovějších aktiv musí být vyšší než výnos aktiv bezrizikových. Na základě přírážky za kreditní riziko zohledněné v cenách dluhopisů (tzv. *reduced-form models*) či akcií (*structural models*) lze odhadnout rizikově neutrální míru selhání dlužníka.

## 2.2.1 Strukturální modely

Zástupcem této skupiny modelů je **Black-Scholes-Mertonův model selhání** (*The Black-Scholes-Merton Default Model*), v jehož rámci dochází k selhání v době splatnosti dluhu v případě, pokud jsou aktiva emitenta nižší než nominální hodnota dluhu.

Jinými slovy pro případ jednoho dluhopisu s nominální hodnotou  $D$  a s dobou splatnosti  $T$  tento model předpokládá selhání dlužníka v čase  $T$  v případě, že  $A_T \leq D$  (kde  $A_T$  je hodnota aktiv v čase  $T$ ). Model určuje hodnotu vlastního kapitálu dlužníka pomocí opční metodologie (Black-Scholesův model pro ocenění call opce), kdy podkladovým aktivem je hodnota aktiv a realizační cenou nominální hodnota dluhu.

Tento model je založen na log-normálním procesu aktiv, v jehož rámci

$$\frac{dA_t}{A_t} = (\mu - \gamma)dt + \sigma dB_t, \quad (2.2)$$

kde  $\mu$  je průměrná míra rentability aktiv,  $\gamma$  je dividendový výnos,  $\sigma$  je volatilita aktiv a  $B$  představuje standardní Brownův pohyb.

Klíčový význam v tomto případě hraje tzv. vzdálenost k úpadku (*distance to default*), která se dá vyjádřit jako

$$X_t = \frac{\log A_t - \log D}{\sigma}, \quad (2.3)$$

kde  $X$  je Brownův pohyb s rozptylem rovným jedné a konstantním přírůstkem

$m = \frac{\left(\mu - \gamma - \frac{\sigma^2}{2}\right)}{\sigma}$ , což je rychlost změny průměrné vzdálenosti k úpadku. Podle

Black-Scholes-Mertonova modelu se úpadek v době zralosti  $T$  vyskytuje s podmíněnou pravděpodobností

$$P(X_T \leq 0 | X_t) = N[u(t, T)], \quad (2.4)$$

kde  $N(x)$  je pravděpodobnost, že proměnná z normovaného normálního rozdělení je menší než  $x$ , a

$$u(t, T) = \frac{X_t + m(T - t)}{\sqrt{T - t}} \quad (2.5)$$

je počet směrodatných odchylek, o který průměrná vzdálenost k úpadku převyšuje nulu v době splatnosti dluhu.

### 2.2.2 Reduced-form modely

Druhým přístupem odhadnutí pravděpodobnosti selhání pomocí tržního ocenění jsou tzv. reduced-form modely. V tomto přístupu dochází k selhání dlužníka, jakmile hodnota jeho aktiv poklesne pod určitou hranici. Na rozdíl od strukturálních modelů k tomuto selhání může dojít kdykoliv.

Reduced-form modely nepovažují vztah mezi úpadkem dlužníka a jeho hodnotou za explicitní. Čas úpadku dlužníka není v těchto modelech určen pomocí jeho celkové hodnoty, ale pomocí prvního skoku exogenně daného skokovým procesem. Taktéž míra návratnosti je u těchto modelů určena exogenně. Parametry, jimiž se řídí míra selhání, jsou odvozeny z tržních dat.

Nejrozšířenějším typem reduced-form modelů jsou tzv. modely intenzity, které jsou založeny na pojmu tzv. intenzity výskytu selhání. Nejjednodušší verze takového modelu definuje úpadek jako čas prvního výskytu selhání  $\tau$  Poissonova procesu s konstantní střední hodnotou míry výskytu, tzv. intenzitou, častou označovanou  $\lambda$ . Potom tedy:

- pravděpodobnost, že dlužník neselže po  $t$  let, je  $p(t) = e^{-\lambda t}$ ;
- předpokládaný čas do úpadku je roven  $1 / \lambda$ ;
- pravděpodobnost úpadku po dobu délky  $\Delta$  je přibližně  $\Delta\lambda$ , pro malé  $\Delta$ .

### 2.2.3 Skóringové modely

Tyto statistické modely patří k nejrozšířenějším nástrojům užívaným k odhadu pravděpodobnosti úpadku dlužníka. Pracují na principu přiřazování vah finančním a ekonomickým ukazatelům, na jejichž základě jsou tyto modely sestaveny. Váhy vyjadřují významnost těchto ukazatelů při odhadování úpadku dlužníka.

Přehled credit scoring modelů je zachycen v následující tabulce (Tab. 2.1).

*Tab. 2.1 Přehled skóringových modelů*

<b>Diskriminační analýza</b>	Lineární diskriminační analýza (LDA)
	Kvadratická diskriminační analýza (QDA)
<b>Regresní modely</b>	Lineární pravděpodobnostní model
	Logit model
	Probit model
<b>Induktivní modely</b>	Neuronové sítě
	Genetické algoritmy

### Lineární diskriminační analýza

Prvním typem skóringových modelů je diskriminační analýza. V případě lineární diskriminační analýzy se jedná o klasifikační techniku, která využívá data z empirického vzorku subjektů k vypočtení lineární diskriminační funkce, která dané subjekty rozděluje do dvou a více skupin na základě jejich charakteristických vlastností.

Parametry diskriminační funkce jsou odhadovány na základě maximalizace rozdílu středních hodnot nezávislých proměnných u všech uvedených skupin, přičemž rozdíly mezi jednotlivými členy stejné skupiny jsou minimalizovány. Rozdíly jsou měřeny pomocí diskriminační proměnné – tzv. *z score*.

Úspěšnost odhadnutého modelu je poté měřena pomocí Wilksovy lambdy, testování modelu a jednotlivých parametrů se provádí na základě F-testu a t-testů.

## **Regresní modely**

Dalším typem skóringových modelů jsou regresní modely, mezi které řadíme lineární pravděpodobnostní model, logit model a probit model. V případě lineárního pravděpodobnostního modelu jsou proměnné vedoucí k úpadku pozorovaných subjektů a jejich váhy identifikovány pomocí prosté lineární regrese. Testování modelu a jednotlivých parametrů se taktéž provádí na základě F-testu a t-testů.

Logit a probit modely jsou vícerozměrné statistické modely sloužící k předpovědi pravděpodobnosti úpadku subjektů, přičemž se jako vstupy využívají různé ekonomické finanční ukazatele. Obecně tyto modely zachycují vztah mezi dichotomickou závislou proměnnou a jednou nebo více nezávislými proměnnými.

Parametry logit a probit modelu jsou odhadovány na základě maximalizace věrohodnostní funkce. K eliminaci parametrů se využívají různé druhy stepwise metody. Parametry modelů se testují převážně pomocí likelihood-ratio testu, Waldova testu a Lagrangova testu.

Podrobnější popis sestavování skóringových modelů pomocí lineární diskriminační analýzy, logit modelu a probit modelu je obsahem třetí kapitoly.

## **Neuronové sítě**

Poslední skupinou skóringových modelů jsou tzv. induktivní modely, mezi něž řadíme neuronové sítě a genetické algoritmy. Zatímco se dříve představené skóringové modely



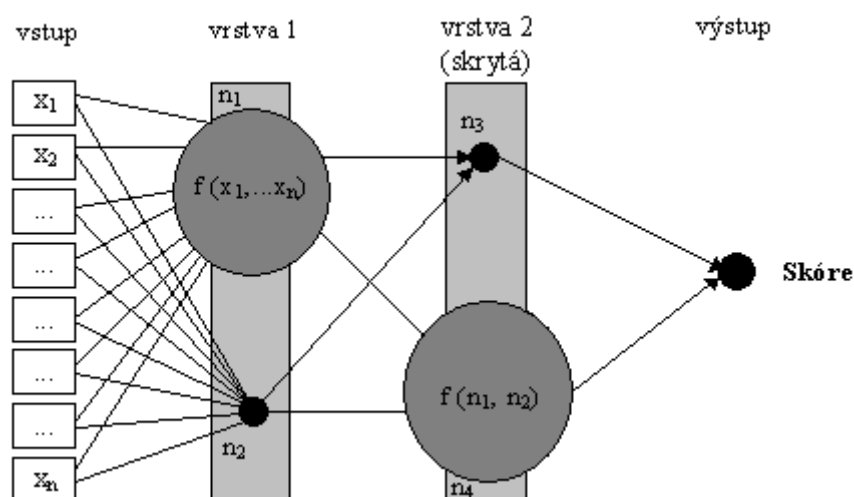
pokoušeli rozpoznat základní vztahy, které by objasnily ekonomicko-finanční rovnováhu subjektu, tyto modely oproti tomu využívají čistě induktivní postup. Pokud je tedy na počátku v datovém vzorku nalezena určitá empirická zákonitost, je tato zákonitost dále použita pro predikci budoucích selhání ostatních subjektů.

Induktivní modely jsou často nazývány jako tzv. „černé skříňky“ (*black boxes*), které se často používají ke generování rychlých výsledků, i když logice postupu nemusí být zcela porozuměno. Tyto černé skříňky také mají jednu podstatnou nevýhodu, kterou je to, že nevíme, co se děje ve skrytých vrstvách mezi vstupy a výstupy. I přesto se však takovýto induktivní postup může ukázat jako velice užitečný, a to v takových případech, kdy je složité nebo téměř nemožné navrhnout odvoditelná pravidla řídící se určitým jevem.

Neuronové sítě se skládají z velkého počtu **neuronů**, což jsou jakési základní prvky. Ty jsou mezi sebou propojeny jednoduchými vazbami nazývanými **synapse**. Neurony jsou uspořádány ve **vrstvách**. Každý neuron v nejvzdálenější vrstvě sítě obdrží jako vstup  $n$  proměnných a zpracuje je buď pomocí lineární, nebo častěji, nelineární funkce. Výsledek se pak předá neuronům v další vrstvě. Tyto neurony taktéž zpracují obdržený vstup pomocí další funkce a odešlou nový výstup další neuronům v následující vrstvě. Po jedné nebo více **skrytých vrstvách** nám síť vygeneruje konečný výsledek. Pokud tedy predikujeme úpadek subjektů, výsledkem může být číselné skóre blížíící se jedničce pro „špatné“ subjekty a blížíící se nule pro ty „zdravé“. Zjednodušené schéma struktury neuronové sítě je znázorněno na obrázku (Obr. 2.1).

Odhad koeficientů jednotlivých funkcí se provádí prostřednictvím opakujícího se mechanismu. Hodnoty těchto koeficientů jsou v praxi tak dlouho upravovány, dokud nezískáme výsledky přibližující se co nejvíce těm, které získat chceme. Z tohoto důvodu je vzdělávací proces sítě postupným pokusem, pomocí něhož se snažíme určit správné váhy proměnných ve vstupu a také synapsí ve skrytých vrstvách.

Obr. 2.1 Zjednodušené schéma struktury neuronové sítě



Zdroj: Resti, A.; Sironi, A. *Risk management and Shareholders' value in banking*, 2007, str. 303

## Genetické algoritmy

Stejně jako neuronové sítě jsou i genetické algoritmy inspirovány chováním živých organismů. Činnost těchto algoritmů je založena na převedení Darwinových principů přirozeného výběru a na konceptu „přežití nejsilnějšího“.

Struktura genetických algoritmů vychází z principů procesu přirozeného vývoje, kdy největší šanci na přežití mají nejsilnější jedinci s nejlepšími vlastnostmi k přizpůsobení se vnějším podmínkám. Tento evoluční proces tak vede k nepřetržitému zdokonalování se druhů. Genetické algoritmy napodobují tento evoluční proces s tím rozdílem, že se nevyvíjejí živé organismy, ale možná řešení problému.

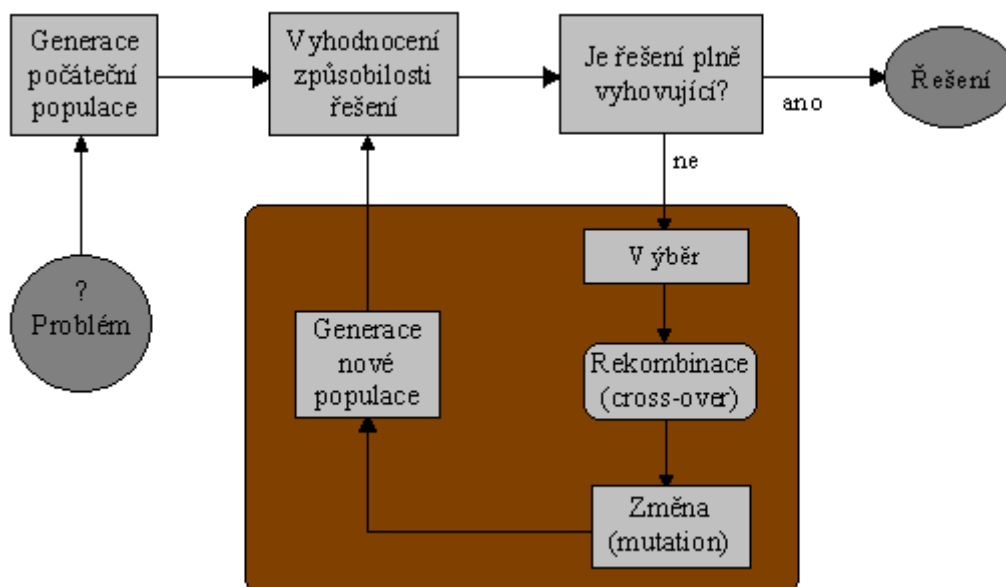
Předpokládejme, že chceme generovat funkci založenou na souboru finančních ukazatelů  $(x_1, x_2, \dots, x_n)$ , která je sestavena takovým způsobem, aby byly přiřazeny nejvyšší hodnoty „zdravým“ subjektům a naopak nízké hodnoty subjektům „špatným“ (v našem případě subjektům v úpadku).

Pokud z důvodu zjednodušení budeme zamýšlet lineární funkci  $z = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m$ , pak bude každé řešení představováno vektorem  $\alpha = [\alpha_0 \dots \alpha_m]'$ , který vyjadřuje algebraický znak a váhu, se kterou jsou různé finanční

ukazatele obsaženy v konstrukci  $z$ . Pokud se u jednoho nebo více  $\alpha_j$  objeví nulové hodnoty, tak to znamená, že odpovídající ekonomicko-finanční ukazatele tímto řešením nejsou použity.

Schéma genetického algoritmu je zobrazeno na následujícím obrázku (Obr. 2.2):

Obr. 2.2 Schéma genetického algoritmu



Zdroj: Resti, A.; Sironi, A. *Risk management and Shareholders' value in banking*, 2007, str. 306

I když genetické algoritmy nezaručují, že identifikují optimální řešení problému, často nám poskytnou dobré řešení a navíc získané velice rychle. Zatímco v určitých oblastech ostatní výzkumné metody zaznamenaly jen chabé výsledky, genetické algoritmy se prokázaly jako mimořádně účinné.

## 2.3 Přehled dnešních komerčních modelů

Na přelomu tisíciletí bylo vyvinuto několik komerčních modelů pro měření kreditního rizika portfolií. Kromě těchto hlavních modelů můžeme ve velkých mezinárodních bankách nalézt různé „interní“ modely, které jsou ve většině případů více či méně inspirovány těmito známými komerčními produkty.

Pro většinu těchto komerčních modelů je snadné nalézt nějakou technickou dokumentaci, která popisuje matematický rámec modelu a dává určitou představu o základních datech a následné kalibraci modelu na tato data. Určitou výjimkou je *PortfolioManager<sup>TM</sup>* model firmy *KMV*, kde je většina dokumentace tajná. Nicméně i u tohoto modelu může být základní myšlenka vysvětlena bez odkazu na neveřejné zdroje.

V následující tabulce (Tab. 2.2) jsou zachyceny čtyři hlavní typy komerčních modelů se společnostmi stojícími za jejich vznikem. Podrobnější popis jednotlivých modelů lze nalézt v Duffie a Singleton (2003), Bluhm, Overbeck a Wagner (2003) či Resti a Sironi (2007).

Tab. 2.2 Přehled dnešních komerčních modelů

<b>Modely hodnoty aktiv</b>	PortfolioManager ( <i>KMV</i> )
	CreditMetrics ( <i>RiskMetrics Group</i> )
<b>Makroekonomické modely</b>	CreditPortfolioView ( <i>McKinsey &amp; Company</i> )
<b>"Actuarian" modely</b>	CreditRisk+ ( <i>Credit Suisse Financial Products</i> )
<b>Modely intenzity</b>	Jarrow/Lando/Turnbull-Model ( <i>Kamakura</i> )
	Duffie/Singleton-Model

### 2.3.1 KMV model (PortfolioManager<sup>TM</sup>)

KMV model známý pod označením *PortfolioManager<sup>TM</sup>* byl vyvinut firmou *KMV* a odhaduje pravděpodobnost úpadku dlužníka na základě Mertonova modelu. S tímto modelem je spojen koncept očekávané frekvence úpadku (**EDF** – *expected default frequencies*). Hlavním výstupem modelu pak je rozložení ztráty úvěrového portfolia.

Model je založen na myšlence, že úpadek či přežití každé firmy závisí na hodnotě aktiv v určitém časovém horizontu. Pokud tato hodnota klesne pod určitou kritickou hranici (v terminologii společnosti *KMV* tzv. *default point*), pak u dlužníka dojde k selhání. Pokud hodnota aktiv zůstane nad touto kritickou hranicí, dlužník „přežívá“.

KMV model má několik výhod, které ho činí populárním, a je tak poměrně hojně užíván bankami po celém světě. První z těchto výhod je rychlost, s jakou se EDF přizpůsobují měnícím se finančním podmínkám společností, které byly hodnoceny. Empirickým

pozorováním bylo zjištěno, že ukazatel EDF reaguje podstatně rychleji na zhoršující se bonitu dlužníka než ratingové agentury, které mají tendenci reagovat s významným zpožděním. Tato výhoda je způsobena tím, že ukazatel EDF je založen na tržních datech.

Další významnou výhodou je to, že ukazatel EDF významně nekolísá z důvodu změny ekonomického cyklu, jak tomu bývá u empirických sazeb selhání spojených s ratingovými agenturami. Obecně mají tyto sazby na rozdíl od EDF tendenci růst v době recese a klesat v době expanze.

Třetí výhoda spočívá v tom, že KMV model přiřazuje každé společnosti specifickou hodnotu ukazatele EDF. Oproti tomu společnosti, kterým je přidělena ratingovými agenturami stejná známka, sdílejí stejnou pravděpodobnost úpadku založenou na historické sazbě selhání pro tuto konkrétní známku. Toto je poměrně významná výhoda vzhledem k tomu, že empirické analýzy ukázaly značnou rozdílnost v pravděpodobnostech úpadku mezi jednotlivými společnostmi, které byly ratingovými agenturami zařazeny do stejné skupiny.

Tento model však má i své nevýhody. Nedá se použít pro výpočet pravděpodobnosti úpadku takových společností, jejichž akcie nejsou obchodovány na burze, neboť tak není známa jejich tržní hodnota a volatilita. Tato skutečnost má dopad na banky, které půjčují především malým a středně velkým podnikům, které nemají přímý přístup na kapitálové trhy.

Další nevýhodou je to, že je tento model založen na předpokladu informačně efektivních akciových trhů. V takovém případě jsou tržní ceny a volatilita výnosů akcií použity k výpočtu tržní hodnoty a volatility aktiv společnosti. Pokud jsou však kapitálové trhy neefektivní, málo likvidní nebo jen neschopné adekvátně promítat všechny dostupné informace do cen, stávají se taková data nespolehlivými.

### **2.3.2 CreditMetrics™**

Jedním z nejznámějších modelů pro odhad kreditního rizika portfolia expozic (ať už úvěrů či dluhopisů) je tzv. *CreditMetrics™* model, který byl poprvé představen americkou bankou *J. P. Morgan*. Tento model je ovšem častěji spojován se společností *RiskMetrics™ Group*, která byla založena právě touto bankou.

*CreditMetrics<sup>TM</sup>* je metoda pro odhad rozložení změn v tržní hodnotě portfolia expozic, které mohou nastat v rámci daného horizontu (většinou jeden rok). Toto rozložení změn může být použito k výpočtu jak očekávané, tak i neočekávané ztráty. *CreditMetrics<sup>TM</sup>* bere v úvahu obě ztráty způsobené selháním a tyto ztráty spojuje s přechodem dlužníka do jiné ratingové třídy, jedná se o tzv. přístup přechodu (*migration approach*).

Ačkoli tento přístup částečně závisí na koncepčních nástrojích vyvinutých Mertonem, jedná se o reduced-form model. Na rozdíl od strukturálních modelů *CreditMetrics<sup>TM</sup>* ve skutečnosti neodvozuje pravděpodobnost úpadku (nebo přechodu) na základě charakteristik společnosti, jakými jsou tržní hodnota a volatilita aktiv či hodnota dluhu, ale používá jako vstupy historické sazby úpadku (nebo přechodu) dané ratingové třídy.

Mezi výhody tohoto modelu patří to, že pracuje s tržními daty, která jsou objektivní, používá přístup tzv. tržní hodnoty, kdy je hodnota úvěrů založena na současné hodnotě budoucích peněžních toků, a také bere v úvahu kromě rizika úpadku i riziko přechodu. Dále je tento model schopen rozpoznat asymetrickou povahu rozložení budoucích hodnot úvěrového portfolia.

Nedostatky modelu *CreditMetrics<sup>TM</sup>* lze spatřovat v tom, že mnoho vstupů potřebných pro tento model lze snadno získat pouze v takovém případě, kdy banka vlastní spolehlivé odhady matic přechodů a existuje likvidní trh, ze kterého lze čerpat informace o forwardových sazbách požadovaných pro různě hodnocené úvěry. Model také předpokládá, že banka operuje na úvěrovém trhu jako příjemce ceny, což znamená, že přírážka za kreditní riziko, jež je určena na základě ratingu dlužníka, je nezávislá na úvěrové politice banky.

Další nevýhoda modelu se týká užívání historických matic přechodů. Historické frekvence totiž ne vždy adekvátně reflektují budoucí pravděpodobnosti, a to ani pokud je matice přechodu v čase stabilní. Navíc je nereálné předpokládat, že jsou tyto pravděpodobnosti stejné pro všechny společnosti patřící do dané ratingové třídy. To souvisí s tím, že *CreditMetrics<sup>TM</sup>* je reduced-form model, který nezohledňuje charakteristiky jednotlivých společností (jejich zadluženost, současnou hodnotu aktiv), ale pracuje pouze s průměrem daným konkrétní ratingovou třídou.

### 2.3.3 CreditPortfolioView™

Tento model, jehož autorem je T. Wilson (pozdější partner v poradenské firmě *McKinsey*), byl vyvinut v roce 1997. *CreditPortfolioView™* je model založený na pozorování, že úvěrové cykly jsou závislé na ekonomickém cyklu. Z tohoto důvodu tedy během fáze hospodářského růstu přechody dlužníků směřujících do vyšších ratingových tříd jsou častější, zatímco pravděpodobnosti přechodu dlužníků do nižších tříd a pravděpodobnost úpadku klesají. Během recese nastává opak. Takto by matice přechodu používané u modelu *CreditMetrics™* měly být upraveny v závislosti na aktuální fázi hospodářského cyklu. *CreditPortfolioView™* proto navrhuje propojit pravděpodobnosti přechodu a úpadku s makroekonomickými ukazateli, jako jsou úroková sazba, míra nezaměstnanosti, růst reálného HDP a sazba úspor.

Snaha o identifikaci vztahu mezi makroekonomickými ukazateli a rizikem úvěrového portfolia je hlavní předností tohoto modelu. V tomto modelu má zhoršení ekonomických podmínek za následek zhoršení hospodářského výhledu dlužníka, stejně jako zvýšení pravděpodobnosti úpadku.

### 2.3.4 CreditRisk+™

*CreditRisk+™* byl v roce 1997 vyvinut firmou *Credit Suisse Financial Products*. Tento model vychází z podobného přístupu, jaký používají pojišťovny. Konkrétně tento přístup vztahuje kreditní riziko k nástrojům typickým pro pojistnou matematiku.

Pojišťovny odvozují své ztráty z frekvence, s jakou se určitý typ události vyskytuje (*event frequency*), a z částky, kterou pojišťovna musí v případě této události vyplatit (*loss severity*). Zde se přímo nabízí analogie s kreditním rizikem, kde ztráty závisí na frekvenci úpadků a na očekávané ztrátovosti v případě selhání dlužníka. Proto je možné použít pojišťovací modely pro odhad kreditních ztrát.

Modely založené na výše popsaném principu jsou zaměřeny pouze na riziko úpadku a neberou v potaz riziko přechodu. Navíc očekávané hodnoty expozice v případě selhání dlužníka a míry návratnosti jsou brány deterministicky, a tak nemůže být odhadnuto ani riziko expozice (*exposure risk*) ani riziko náhrady (*recovery risk*).

Navzdory těmto omezením je *CreditRisk<sup>+</sup>* model vysoce efektivní v odhadování kreditního rizika portfolií s velkým počtem pozic, a je tak hojně využíván bankami při řízení tradičních portfolií, jako jsou úvěry poskytnuté malým a středním podnikům, spotřebitelské půjčky či hypotéky.

Na rozdíl od strukturálních modelů tento model nevysvětluje proces vedoucí k úpadku dlužníka a ani neodhaduje pravděpodobnost úpadku. Naopak předpokládá, že pravděpodobnosti úpadku dlužníků a míry návratnosti jejich půjček již byly odhadnuty jinými nástroji (pomocí interního ratingového systému banky nebo studiem externích ratingových agentur).



## 3 Sestavení skóringových modelů

Jak již bylo zmíněno v předcházející kapitole, skóringové modely (*credit scoring models*) jsou hojně využívány pro odhad pravděpodobnosti úpadku dlužníka. V této kapitole budou popsány tři typy skóringových modelů, konkrétně logit a probit model, jež se řadí mezi regresní modely, a dále lineární diskriminační analýza. Tyto tři modely vysvětlují stav finančního zdraví dlužníka na základě volby vhodných proměnných, často na bázi struktury finančních výkazů.

### 3.1 Regresní modely

Regresní metody se staly nedílnou součástí analýzy dat, jejichž cílem je popsat vztah mezi závislou (vysvětlovanou) proměnnou a jednou nebo více nezávislými (vysvětlujícími) proměnnými. V praxi se často setkáváme s případem, kdy je závisle proměnná diskrétní, tzn. může nabývat dvou (v našem případě „*default*“ a „*non-default*“) a více možných hodnot. Logistická i probit regrese se staly v mnoha oblastech standardní metodou k analýze takovéto situace. Následující podkapitoly týkající se logistické regrese budou vycházet především z publikace Hosmer a Lemeshow (2000).

#### 3.1.1 Logit model

Mezi lineární a logistickou regresí existují dva významné rozdíly. První rozdíl souvisí s povahou vztahu mezi závislou a nezávislou proměnnou. V případě jakéhokoliv problému, na nějž je aplikována regresní metoda, je klíčovou veličinou střední hodnota závisle proměnné daná hodnotou nezávisle proměnné. Tato veličina se nazývá **podmíněná střední hodnota** (*conditional mean*) a bývá vyjádřena jako „ $E(Y|x)$ “, kde  $Y$  označuje závisle proměnnou a  $x$  hodnotu nezávisle proměnné.

Pokud budeme uvažovat souhrn  $p$  nezávislých proměnných daný vektorem  $x' = (x_1, x_2, \dots, x_p)$ , pak se tato střední hodnota v lineární regresi vypočte dle následujícího vztahu:

$$E(Y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p, \quad (3.1)$$

kde  $\beta_0$  značí úrovnovou konstantu a  $\beta_j$  (pro  $j = 1, 2, \dots, p$ ) značí koeficienty pro každou z nezávislých proměnných. Dle tohoto výrazu tedy tato střední hodnota může nabývat jakýchkoliv hodnot (od  $-\infty$  po  $+\infty$ ).

Existují dva hlavní důvody proč v případě binární (dichotomické) závisle proměnné použít logistickou distribuční funkci. Prvním z těchto důvodů je to, že z matematického hlediska je tato funkce velice flexibilní a snadno použitelná. Druhým důvodem je potom možnost její smysluplné interpretace.

V případě logistické regrese označme podmíněnou střední hodnotu  $E(Y|x)$  jako  $\pi(x)$ . Poté můžeme logit model, kde se používá tzv. **logistická transformace** (*logit transformation*), vyjádřit jako:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}. \quad (3.2)$$

Tímto vztahem je zajištěno, že podmíněná střední hodnota může nabývat hodnot pouze od nuly do jedné. Pokud vezmeme v úvahu, že podmíněná střední hodnota  $\pi(x)$  je v našem případě pravděpodobnost úpadku (**PD** – *probability of default*) dlužníka, pak po úpravě vztahu (3.2) dostáváme:

$$\pi(x) = PD = \frac{1}{1 + e^{-g(x)}}, \quad (3.3)$$

kde nezávisle proměnnou  $g(x)$  pro souhrn  $p$  nezávislých proměnných daný vektorem  $x' = (x_1, x_2, \dots, x_p)$  vypočteme dle následujícího vztahu:

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p. \quad (3.4)$$

Dosazením (3.4) do (3.3) získáme konečnou podobu logit modelu:

$$\pi(x) = PD = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}. \quad (3.5)$$

Druhý důležitý rozdíl mezi logistickou a lineární regresí se týká podmíněného rozdělení závisle proměnné. V případě lineární regrese předpokládáme, že pozorování závisle proměnné mohou být vyjádřena jako  $y = E(Y|x) + \varepsilon$ , kde  $\varepsilon$  značí reziduální odchylku, jež vyjadřuje odchylku od podmíněné střední hodnoty. Nejčastěji se předpokládá, že  $\varepsilon$  má normální rozdělení s nulovou střední hodnotou a s konstantním rozptylem, z čehož vyplývá, že podmíněné rozdělení závisle proměnné dané  $x$  bude také normální se střední hodnotou  $E(Y|x)$  a s konstantním rozptylem.

V případě logistické regrese můžeme vyjádřit hodnotu závisle proměnné danou  $x$  jako  $y = \pi(x) + \varepsilon$ . Zde reziduální odchylka  $\varepsilon$  může nabývat dvou různých hodnot. Pokud  $y = 1$ , potom  $\varepsilon = 1 - \pi(x)$  s pravděpodobností  $\pi(x)$ , a pokud  $y = 0$ , pak  $\varepsilon = -\pi(x)$  s pravděpodobností  $1 - \pi(x)$ . Reziduální odchylka  $\varepsilon$  má tedy v tomto případě rozdělení s nulovou střední hodnotou a s rozptylem rovným  $\pi(x)[1 - \pi(x)]$ . Podmíněné rozdělení závisle proměnné je tedy binomické rozdělení s pravděpodobností danou podmíněnou střední hodnotou  $\pi(x)$ .

### 3.1.2 Probit model

Druhým regresním modelem, který se používá v případech, kdy závisle proměnná nabývá dvou a více možných hodnot, je tzv. probit model. U tohoto modelu se na rozdíl od logistické transformace použité u logit modelu aplikuje **kumulativní distribuční funkce normovaného normálního rozdělení**.

V našem případě je podmíněná střední hodnota opět rovna pravděpodobnosti úpadku, tedy  $y = PD = E(Y|x)$ . Pokud budeme uvažovat souhrn  $p$  nezávislých proměnných daný vektorem  $x' = (x_1, x_2, \dots, x_p)$ , můžeme probit model vyjádřit následovně:

$$y = PD = \int_{-\infty}^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt \quad (3.6)$$

neboli

$$y = PD = \Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p), \quad (3.7)$$

kde  $\Phi$  značí kumulativní distribuční funkci normovaného normálního rozdělení.

Stejně jako v případě logistické regrese se i probit regrese od lineární regrese liší v povaze vztahu mezi závisle a nezávisle proměnnou a podmíněným rozdělením závisle proměnné.

Logit a probit model si jsou velice podobné. Hlavním rozdílem je to, že logistická funkce má těžší konce (tzv. *fatter tails*).

### **3.2 Odhad parametrů a testování statistické významnosti logit a probit modelu**

V následujících podkapitolách budou charakterizovány postupy, jež jsou pro logit i probit model totožné.<sup>1</sup> Konkrétně se bude jednat o odhad parametrů, testování významnosti parametrů, odhad konfidenčního intervalu a taktéž bude popsána stepwise metoda pro výběr parametrů.

#### **3.2.1 Odhad parametrů**

Předpokládejme, že máme vzorek  $n$  nezávislých pozorování  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , kde  $y_i$  značí hodnotu závisle proměnné a  $x_i$  hodnotu nezávisle proměnné pro  $i$ -tý subjekt. Dále předpokládejme, že závisle proměnná může nabývat pouze dvou možných hodnot (v našem případě 0 pro „*non-default*“ a 1 pro „*default*“). Abychom podle vztahů (3.5) a (3.7) mohli modely sestavit, musíme odhadnout hodnoty vektoru  $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ , tzv. parametry.

---

<sup>1</sup> Vzorce, které obsahují podmíněnou střední hodnotu – (3.8) až (3.11), jsou uvedeny pro logit model. V případě probit modelu by se změnilo pouze značení této střední hodnoty.

Na rozdíl od lineární regrese, kde se pro odhad parametrů používá **metoda nejmenších čtverců** (OLS – *ordinary least squares method*), se v logistické a probit regresi používá tzv. **metoda maximální věrohodnosti** (*maximum likelihood method*). Tato metoda používá k nalezení nejlepšího odhadu neznámých parametrů **věrohodnostní funkci** (*likelihood function*). Nejlepší odhad pak získáme pro takovou hodnotu parametru, pro kterou dosahuje věrohodnostní funkce svého maxima.

Za předpokladu, že pozorování jsou nezávislá, věrohodnostní funkce vypadá následovně:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}. \quad (3.8)$$

V praxi se však častěji používá logaritmus věrohodnostní funkce, neboť se s ním lépe počítá. Logaritmus této funkce je tedy definován jako:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\}. \quad (3.9)$$

Abychom našli hodnoty vektoru  $\beta'$ , které maximalizují věrohodnostní funkci  $L(\beta)$ , derivujeme tuto funkci s ohledem na  $p + 1$  parametrů a výsledné výrazy položíme rovny nule. Takto získáme následující rovnice (tzv. *likelihood equations*):

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (3.10)$$

a

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \quad (3.11)$$

pro  $j = 1, 2, \dots, p$ .

Vzhledem k tomu, že výrazy ve vzorcích (3.10) a (3.11) jsou nelineární v neznámých parametrech, řešení těchto rovnic vyžaduje speciální metody. Tyto metody jsou součástí většiny statistických programů.

Hodnoty vektoru  $\beta'$  dané řešeními rovnic (3.10) a (3.11) jsou nazývány jako **odhady maximální věrohodnosti** (*maximum likelihood estimates*) a označují se jako  $\hat{\beta}$ . Potom predikované hodnoty logit modelu nesou označení  $\hat{\pi}(x_i)$ , což jsou hodnoty vypočtené dle vztahu (3.5) pomocí  $\hat{\beta}$  a  $x_i$ . V případě probit modelu predikované hodnoty nesou označení  $\hat{y}_i$  a jsou vypočteny dle vztahu (3.7).

### 3.2.2 Testování významnosti parametrů

Pro testování statistické významnosti parametrů logit a probit modelu se dají použít tři testy: likelihood-ratio test, Waldův test a Lagrangův test (score test). Tyto testy jsou někdy popisovány jako testy rozdílů mezi dvěma modely, kdy je vždy jeden z modelů součástí dalšího modelu (tzv. *nested models*).

#### Likelihood-ratio test

Likelihood-ratio test se provádí odhadem dvou modelů a porovnáním těchto dvou modelů mezi sebou. Odstranění určitých nezávislých proměnných z modelu má téměř vždy za následek zhoršení jeho vypovídací schopnosti, ovšem je nutné sledovat, zda tento pozorovaný rozdíl je statisticky významný. Likelihood-ratio test tohle provádí porovnáním **logaritmů věrohodnostních funkcí** těchto dvou modelů získaných ze vztahu (3.9), tzv. *log-likelihoods*. Pokud je tento rozdíl statisticky významný, pak méně restriktivní model (ten s více proměnnými) má větší vypovídací schopnost než více restriktivní model.

Likelihood-ratio statistiku lze vypočítat dle následujícího jednoduchého vztahu:

$$LR = -2(L_0 - L_1), \quad (3.12)$$

kde  $L_0$  značí log-likelihood hodnotu více restriktivního modelu a  $L_1$  méně restriktivního modelu.

Výsledná statistika testu má  $\chi^2$  (*chi-square*) rozdělení se stupni volnosti rovnajícím se počtu proměnných, které byly z modelu vyloučeny.

## Waldův test

Waldův test má jednu výhodu oproti likelihood-ratio testu a to tu, že vyžaduje odhadnutí pouze jednoho modelu. Tento test pracuje s nulovou hypotézou, že určitý parametr nebo soubor parametrů jsou rovny nule. Pokud test tuto nulovou hypotézu nezamítne, značí to, že odstraněním proměnné nebo souboru proměnných nedojde k výraznému snížení vypovídací schopnosti modelu, jelikož proměnná s velmi nízkou hodnotou parametru ve srovnání s jeho směrodatnou odchylkou obecně příliš nepřispívá k predikci závisle proměnné.

Vztah pro výpočet Waldovy statistiky poměří hodnotu odhadnutého parametru s jeho směrodatnou odchylkou (**SE** – *standard error*).

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (3.13)$$

Další výhoda tohoto testu spočívá v tom, že můžeme testovat více parametrů současně.

## Lagrangův (score) test

Na rozdíl od prvních dvou testů tento test nevyžaduje odhady maximální věrohodnosti pro parametry, což je považováno za jeho největší výhodu. Tento test můžeme použít k testování toho, zda přidání dalších proměnných do modelu bude mít za následek významné zlepšení vypovídací schopnosti modelu.

Statistika testu se vypočítá na základě sklonu věrohodnostní funkce pozorovaných hodnot proměnných v modelu. Odhadnutý sklon, někdy také označován jako „skóre“, je důvodem, proč se někdy Lagrange test označuje právě jako **score test**. Tato skóre jsou následně použita pro odhad zlepšení vypovídací schopnosti modelu, pokud do modelu byly přidány další proměnné.

### 3.2.3 Odhad konfidenčního intervalu

Výpočet konfidenčních intervalů (jinak taky intervalů spolehlivosti) pro jednotlivé parametry je důležitým doplňkem testování významnosti odhadnutých modelů. Základem pro konstrukci intervalových odhadů je stejná statistická teorie, která byla použita při testování významnosti modelů. Cílem je určení intervalu, který s předem zvolenou pravděpodobností zahrnuje hodnotu neznámého parametru.

Koncové body  $100(1 - \alpha)\%$  konfidenčního intervalu, kde  $\alpha$  je zvolená hladina významnosti, pro úrovnovou konstantu se vypočtou dle následujícího vztahu:

$$\hat{\beta}_0 \pm z_{1-\alpha/2} SE(\hat{\beta}_0), \quad (3.14)$$

a pro parametry nezávislých proměnných dle vztahu:

$$\hat{\beta}_j \pm z_{1-\alpha/2} SE(\hat{\beta}_j) \quad (3.15)$$

pro  $j = 1, 2, \dots, p$ , kde  $z_{1-\alpha/2}$  je horní kritická hodnota z normovaného normálního rozdělení na hladině významnosti  $\alpha/2$  a  $SE(\cdot)$  značí odhad směrodatné odchylky příslušného parametru.

### 3.2.4 Stepwise metoda

Stepwise metoda je algoritmus pro odhad regresního modelu a používá se v takovém případě, kdy máme k dispozici velké množství nezávislých proměnných, jejichž počet chceme snížit natolik, aby v konečném modelu zůstaly pouze ty významné. Používaná statistika u stepwise metody závisí na předpokladech modelu. Zatímco v případě lineární regrese se používá  $F$ -test, jelikož reziduální odchylky mají normální rozdělení, v logistické a probit regresi mají náhodné chyby binomické rozdělení a statistická významnost proměnné se tedy většinou posuzuje pomocí likelihood-ratio testu. V každé fázi tohoto postupu je



„nejdůležitější“ proměnná ta, která vyvolá největší změnu v log-likelihood hodnotě vzhledem k modelu, který tuto proměnnou neobsahuje.

Hlavními postupy stepwise metody jsou:

- **forward selection** – začíná se s „prázdným“ modelem (model obsahující pouze konstantu), proměnné se přidávají jedna po druhé a testuje se, zda jsou statisticky významné;
- **backward elimination** – začíná se s modelem obsahujícím všechny proměnné, jedna po druhé se testují a z modelu se vyřazují ty, které jsou statisticky nevýznamné;
- metody, které jsou kombinací výše uvedených metod, kdy se v každé fázi postupu proměnné testují, zda mají být do modelu zahrnuty nebo vyloučeny.

Postup stepwise metody (*forward selection* následovaná *backward elimination*) v případě logistické a probit regrese můžeme rozdělit do několika kroků.

## Krok (0)

Předpokládejme, že máme celkem  $p$  nezávislých proměnných, které jsou všechny považovány za důležité při studování závisle proměnné. Krok (0) začíná odhadnutím modelu obsahujícím pouze konstantu a výpočtem jeho log-likelihood,  $L_0$ . Následuje odhad každého z  $p$  možných regresních modelů (modely obsahující konstantu a jednu nezávisle proměnnou) a srovnání jejich příslušných log-likelihoods. Necht' je hodnota pro log-likelihood modelu obsahujícího proměnnou  $x_j$  v kroku nula označena jako  $L_j^{(0)}$ . Dolní index „ $j$ “ odkazuje na proměnnou, která byla do modelu přidána, a horní index „(0)“ odkazuje na číslo kroku. Toto značení bude dodržováno i nadále.

Necht' je hodnota likelihood-ratio testu pro model obsahující proměnnou  $x_j$  a pro model obsahující pouze konstantu označena jako  $LR_j^{(0)} = -2(L_0 - L_j^{(0)})$  a  $p$ -hodnota tohoto testu označena jako  $p_j^{(0)}$ . Tato  $p$ -hodnota je stanovena pravděpodobností  $\Pr[\chi^2(\nu) > G_j^{(0)}] = p_j^{(0)}$ , kde  $\nu=1$ , pokud je proměnná  $x_j$  spojitá.

Nejvýznamnější proměnnou je ta s nejmenší  $p$ -hodnotou. Pokud tuto proměnnou označíme jako  $x_{e_1}$ , potom  $p_{e_1}^{(0)} = \min(p_j^{(0)})$ . Dolní index „ $e_1$ “ označuje proměnnou, která je kandidátem pro vstup do modelu v kroku (1). Ovšem jen proto, že  $x_{e_1}$  je nejvýznamnější proměnnou, neznamená to ještě, že je statisticky významná.

Klíčovým aspektem při použití stepwise metody je volba tzv. hladiny významnosti pro přidání proměnné do modelu. Necht' je  $p_E$  označení pro tuto volbu, kde „ $E$ “ značí vstup do modelu (*entry*). Volba  $p_E$  určuje, kolik proměnných bude nakonec do modelu zahrnuto. Ať už je  $p_E$  stanoveno jakkoliv, proměnná je považována za dost významnou pro zahrnutí do modelu v takovém případě, pokud je  $p$ -hodnota LR menší než  $p_E$ . Dále přecházíme ke kroku (1), pokud  $p_{e_1}^{(0)} < p_E$ ; v opačném případě tento proces ukončíme.

## Krok (1)

Krok (1) začíná odhadnutím modelu obsahujícím proměnnou  $x_{e_1}$ . Necht' je  $L_{e_1}^{(1)}$  označení log-likelihood hodnoty tohoto modelu. K určení toho, zdali je nějaká z  $p - 1$  zbývajících proměnných významná, jakmile již proměnná  $x_{e_1}$  je v modelu, odhadujeme  $p - 1$  regresních modelů obsahujících tuto proměnnou a  $x_j$  ( $j = 1, 2, 3, \dots, p$  a  $j \neq e_1$ ). Pro model obsahující proměnné  $x_{e_1}$  a  $x_j$  necht' je log-likelihood hodnota označena jako  $L_{e_1j}^{(1)}$  a hodnota likelihood-ratio testu tohoto modelu a modelu obsahujícího pouze proměnnou  $x_{e_1}$  označena jako  $LR_j^{(1)} = -2(L_{e_1}^{(1)} - L_{e_1j}^{(1)})$ . Potom  $p$ -hodnota tohoto testu je označena jako  $p_j^{(1)}$ . Necht' je proměnná s nejmenší  $p$ -hodnotou v kroku (1) označena jako  $x_{e_2}$ , kde  $p_{e_2}^{(1)} = \min(p_j^{(1)})$ . Pokud je tato hodnota menší než  $p_E$ , přecházíme ke kroku (2); v opačném případě proces ukončíme.

## Krok (2)

Krok (2) začíná odhadnutím modelu obsahujícím proměnné  $x_{e_1}$  a  $x_{e_2}$ . Ovšem existuje zde možnost, že jakmile byla proměnná  $x_{e_2}$  přidána do modelu, proměnná  $x_{e_1}$  již není nadále významná. Proto krok (2) zahrnuje i zpětnou eliminaci proměnných (*backward elimination*). Ta se obecně provádí odhadnutím modelů, ze kterých se vyloučí jedna z proměnných přidávaných v předchozích krocích, a hodnocením významnosti této vyloučené proměnné.

Nechť je v kroku (2)  $L_{-e_j}^{(2)}$  označení pro log-likelihood hodnotu modelu, ze kterého byla vyloučena proměnná  $x_{e_j}$ . Podobně nechť je likelihood-ratio test tohoto modelu a „plného“ modelu v kroku (2) označen jako  $LR_{-e_j}^{(2)} = -2(L_{-e_j}^{(2)} - L_{e_1 e_2}^{(2)})$  a jeho  $p$ -hodnota jako  $p_{-e_j}^{(2)}$ .

Abychom zjistili, zdali by proměnná měla být vyloučena z modelu, musíme vybrat proměnnou, která pokud je vyloučena, tak nám dá maximální  $p$ -hodnotu. Pokud tuto proměnnou tedy označíme jako  $x_{r_2}$ , potom  $p_{r_2}^{(2)} = \max(p_{-e_j}^{(2)}, p_{-e_2}^{(2)})$ . Abychom rozhodli, zdali proměnná  $x_{r_2}$  má být z modelu vyloučena, musíme porovnat hodnotu  $p_{r_2}^{(2)}$  s druhou hladinou významnosti, tentokrát pro vyloučení proměnné z modelu. Tuto hladinu označme jako  $p_R$ , kde „R“ značí vyloučení z modelu (*remove*). Ať už zvolíme jakoukoliv hodnotu  $p_R$ , musí být tato hodnota větší než hodnota  $p_E$ , abychom vyloučili možnost, že bude stejná proměnná zařazena a následně vyloučena z modelu v po sobě jdoucích krocích. Pokud je hodnota  $p_{r_2}^{(2)}$  větší než hodnota  $p_R$ , potom je proměnná  $x_{r_2}$  z modelu vyloučena. V opačném případě tato proměnná v modelu zůstává. V obou případech následuje fáze výběru dalších proměnných.

V této fázi je odhadován každý z  $p - 2$  regresních modelů obsahujících proměnné  $x_{e_1}$ ,  $x_{e_2}$  a  $x_j$  ( $j = 1, 2, 3, \dots, p$  a  $j \neq e_1, e_2$ ). Vypočítají se log-likelihood hodnoty pro každý model, provedou se likelihood-ratio testy porovnávající každý z těchto modelů s modelem obsahujícím pouze proměnné  $x_{e_1}$  a  $x_{e_2}$  a určí se příslušné  $p$ -hodnoty těchto testů. Nechť je  $x_{e_3}$  označení pro proměnnou s nejmenší  $p$ -hodnotou, potom tedy  $p_{e_3}^{(2)} = \min(p_j^{(2)})$ . Pokud  $p_{e_3}^{(2)} < p_E$ , přecházíme ke kroku (3); v opačném případě proces ukončíme.

### Krok (3)

V tomto kroku je postup stejný jako v kroku (2). Odhaduje se model, který již obsahuje i proměnnou vybranou v předcházejícím kroku, provádí se zpětná eliminace proměnných následovaná fází výběru dalších proměnných. Tímto způsobem proces pokračuje až do posledního kroku, krok (S).

### Krok (S)

Tento krok nastane v případě, pokud:

- všech  $p$  proměnných je zařazeno do modelu;
- všechny proměnné v modelu mají  $p$ -hodnoty menší než  $p_R$  a proměnné, jež do modelu zařazeny nebyly, mají  $p$ -hodnoty větší než  $p_E$ .

V tomto kroku jsou do modelu zařazeny takové proměnné, které jsou důležité vzhledem ke kritériím  $p_E$  a  $p_R$ .

### 3.3 Lineární diskriminační analýza

Další z možností, jak sestavit skóringový model, je tzv. lineární diskriminační analýza. Cílem této analýzy je najít takové nezávisle proměnné, které by byly významné pro rozlišení studovaných prvků. V kontextu kreditního rizika se jedná o určení, zda je dlužník finančně zdravý či nikoliv. To zjišťujeme na základě údajů zjištěných z jednotlivých finančních výkazů. Jako kritérium pro úpadek subjektu může být považováno výskyt insolvenčního řízení, finanční restrukturalizace či označení úvěrů jako pochybných.

Uvažujme počet  $n$  dlužníků,  $n = 1, \dots, N$ , které podle aktuální situace rozdělíme do dvou skupin; „dobré“ dlužníky do **skupiny A** a dlužníky v úpadku do **skupiny B**. Pro všechny dlužníky známe hodnoty dvou ukazatelů  $x_1$  a  $x_2$ . Úkolem lineární diskriminační analýzy je přidělit těmto dvěma ukazatelům takový koeficient  $\gamma$ , aby celkové skóre  $z$  vypočtené dle vztahu

$$z_n = \gamma_1 x_{1,n} + \gamma_2 x_{2,n} \quad (3.16)$$

umožnilo co nejlepší rozdělení dobrých dlužníků od těch v úpadku. Pokud bychom tento vztah zobecnili pro celkový počet  $m$  ukazatelů, mohli bychom jej přepsat následovně:

$$z_i = \sum_{j=1}^m \gamma_j x_{j,i}. \quad (3.17)$$

Rozčlenění dobrých dlužníků od těch v úpadku je prováděno na základě maximalizace rozdílu mezi středními hodnotami skóre pro skupiny A a B,

$$z_A - z_B \rightarrow \max, \quad (3.18)$$

kdy se snažíme minimalizovat odchylku ve skóre dlužníků uvnitř obou skupin. Vektor koeficientů  $\gamma$  je tedy možné určit následovně:

$$(\gamma_1, \gamma_2) = \sum^{-1} (x_1^A - x_1^B, x_2^A - x_2^B), \quad (3.19)$$

přičemž,

$$\sum = \frac{n_A - 1}{n_A + n_B - 2} \sum_A + \frac{n_B - 1}{n_A + n_B - 2} \sum_B, \quad (3.20)$$

kde  $\sum_A$  ( $\sum_B$ ) je kovarianční matice ukazatelů dobrých dlužníků, resp. dlužníků v úpadku.

### 3.3.1 Statistická verifikace odhadnutých parametrů a modelu

Statistická verifikace je zkoumána pomocí tzv.  $t$ -testů a  $F$ -testu. Zatímco  $t$ -testy slouží ke stanovení statistické významnosti odhadnutých koeficientů,  $F$ -test slouží pro posouzení významnosti jako celku. Popis obou těchto testů bude vycházet z publikace Zmeškal a kol. (2004).

#### **$t$ -test**

Tento test je využíván pro testování hypotézy o individuálních regresních koeficientech. Nulová hypotéza předpokládá nevýznamnost koeficientů, naopak alternativní hypotéza předpokládá, že koeficienty jsou statisticky významné.

$$H_0: \hat{\beta}_i = 0,$$

$$H_A: \hat{\beta}_i \neq 0.$$

Test je prováděn pomocí ***t*-statistiky**, přičemž se předpokládá, že tato statistika má Studentovo rozdělení pravděpodobnosti s  $df$ -stupni volnosti,

$$t_{df} = \frac{\hat{\beta}_i - 0}{SE_{\hat{\beta}_i}}, \quad (3.21)$$

kde  $SE_{\hat{\beta}_i}$  je odhad směrodatné odchylky koeficientu  $\hat{\beta}_i$ . Vyhodnocovací pravidlo je založeno na porovnání dvou parametrů:  $t^{vyp}$  ( $t$ -statistice vypočtené, odpovídající dané odhadované hodnotě  $\hat{\beta}_i$ ) a  $t^{krit}$  ( $t$ -kritické, určující percentil  $t$ -statistiky na dané úrovni významnosti  $\alpha$ ),

$$t_{df}^{vyp} = \frac{\hat{\beta}_i}{SE_{\hat{\beta}_i}}, \quad (3.22)$$

$$t_{\alpha/2;df}^{krit} = ST_{df}^{-1}(\alpha/2), \quad (3.23)$$

kde  $ST$  je distribuční funkce Studentova rozdělení a  $ST_{\alpha/2;df}^{-1}$  potom inverzní funkce (kvantil) na hladině pravděpodobnosti  $\alpha/2$  s  $df$  stupni volnosti. Oboustrannou pravděpodobnost dosažení hodnoty  $t^{vyp}$  vypočteme dle následujícího vztahu:

$$hodnota P_{df} = \alpha^{vyp} = ST_{df}(t_{df}^{vyp}) \cdot 2. \quad (3.24)$$

Rozhodovací pravidlo pro oboustranný test lze formulovat následovně. Zamítnutí nulové hypotézy (a tedy přijetí alternativní) lze stanovit dvěma způsoby:

- a) jestliže  $|t_{df}^{vyp}| > t_{\alpha/2;df}^{krit}$ ,
- b) jestliže  $hodnota P_{df} < \alpha$ .

Přijetí nulové hypotézy (a tedy zamítnutí alternativní) lze rovněž určit dvěma způsoby:

- a) jestliže  $|t_{df}^{vyp}| \leq t_{\alpha/2;df}^{krit}$ ,

b) jestliže  $hodnota P_{df} \geq \alpha$ .

Zamítnutí nulové hypotézy znamená, že propočtený koeficient leží v kritické oblasti, je statisticky významný a ze statistického pohledu má být zařazen do odhadovaného modelu. Při přijetí nulové hypotézy pak platí opak.

## **F-test**

Pro stanovení významnosti modelu se pak používá  $F$ -test, který je založen na podobných principech jako předcházející  $t$ -test. Nulová hypotéza je stanovena na předpokladu, že všechny koeficienty jsou nulové. Naopak alternativní hypotéza tvrdí, že alespoň jeden koeficient je různý od nuly. Pokud budeme uvažovat souhrn  $p$  nezávislých proměnných daný vektorem  $x' = (x_1, x_2, \dots, x_p)$ , potom bude matematické vyjádření těchto hypotéz následující:

$$H_0: \hat{\beta}_i = 0,$$

$$H_A: \hat{\beta}_0 \neq 0 \vee \hat{\beta}_1 \neq 0 \vee \dots \vee \hat{\beta}_p \neq 0 \text{ (alespoň jeden z parametrů je simultánně nenulový).}$$

Tento test je konstruován na bázi **F-statistiky** za předpokladu, že tato statistika má Fischerovo rozdělení pravděpodobnosti,

$$F = \frac{ESS / df_{ESS}}{RSS / df_{RSS}} = \frac{MS_{ESS}}{MS_{RSS}}, \quad (3.25)$$

kde  $ESS$  je rozptyl vysvětlený modelem (*explained sum of squares*),  $RSS$  je rozptyl přiřazen reziduálnímu rozptylu nevysvětlenému modelem (*residual sum of squares*),  $MS_{ESS}$  je průměrný vysvětlený rozptyl (*mean square explained*),  $MS_{RSS}$  je průměrný reziduální rozptyl (*mean square residual*),  $df_{ESS}$  a  $df_{RSS}$  jsou stupně volnosti přiřazené uvedeným rozptylům,  $df_{ESS} = k + 1$  a  $df_{RSS} = T - (k + 1)$ , kde  $k$  je počet nezávislých proměnných a  $T$  je počet pozorování. Jednička je v těchto vzorcích přičítána pouze v případech, pokud je v modelu zahrnuta i úrovněová konstanta.

Vyhodnocení je založeno na porovnání hodnoty vypočtené ( $F^{vyp}$ ) a kritické ( $F^{krit}$ ) statistiky. Vychází se z předpokladu, že  $F$ -statistika má Fischerovo rozdělení pravděpodobnosti,

$$F_{df_{ESS}; df_{RSS}}^{vyp} = \frac{MS_{ESS}}{MS_{RSS}}, \quad (3.26)$$

$$F_{\alpha; df_{ESS}; df_{RSS}}^{krit} = FISH_{df_{ESS}; df_{RSS}}^{-1}(\alpha), \quad (3.27)$$

kde  $FISH$  je distribuční funkce Fisherova rozdělení,  $FISH_{df_{ESS}; df_{RSS}}^{-1}$  je inverzní funkce (kvantil) na hladině pravděpodobnosti  $\alpha$ .

$$hodnota P_{df_{ESS}; df_{RSS}} = \alpha^{vyp} = FISH_{df_{ESS}; df_{RSS}}(F^{vyp}). \quad (3.28)$$

Rozhodovací pravidlo pro jednostranný  $F$ -test lze formulovat následovně. Zamítnutí nulové hypotézy (a tedy přijetí alternativní) lze stanovit dvěma způsoby:

- a) jestliže  $F_{df_{ESS}; df_{RSS}}^{vyp} > F_{\alpha; df_{ESS}; df_{RSS}}^{krit}$ ,
- b) jestliže  $hodnota P_{df_{ESS}; df_{RSS}} < \alpha$ .

Přijetí nulové hypotézy (a tedy zamítnutí alternativní) lze rovněž určit dvěma způsoby:

- a) jestliže  $F_{df_{ESS}; df_{RSS}}^{vyp} \leq F_{\alpha; df_{ESS}; df_{RSS}}^{krit}$ ,
- b) jestliže  $hodnota P_{df_{ESS}; df_{RSS}} \geq \alpha$ .

Zamítnutí nulové hypotézy znamená, že odhadnutý model je jako celek statisticky významný.



### 3.3.2 Míra úspěšnosti modelu

Míra úspěšnosti modelu slouží k určení, nakolik je model sestavený pomocí diskriminační analýzy úspěšný v oddělení dobrých dlužníků od těch, kteří jsou v úpadku. Nejdříve je třeba vypočíst tzv. **Wilksovo lambda**, a to jako podíl součtu odchylek obou skupin na celkových odchylkách:

$$\Lambda = \frac{\sum_{i \in A} (z_i - z_A)^2 + \sum_{i \in B} (z_i - z_B)^2}{\sum_{i=1}^n (z_i - \bar{z})^2}, \quad (3.29)$$

kde  $\bar{z}$  představuje střední hodnotu  $z_i$  v celém vzorku dlužníků (dobrých i těch v úpadku).

K explicitnímu vyjádření míry úspěšnosti modelu je třeba odečíst hodnotu Wilksovy lambdy od jedničky, tedy  $1 - \Lambda$ .

### 3.3.3 Přepočet skóre na pravděpodobnost úpadku

I když může být lineární diskriminační analýza rovněž použita k přímému odhadu pravděpodobnosti úpadku jednotlivých dlužníků, tak na rozdíl od logit a probit modelů je nutné tuto pravděpodobnost přepočítat z již vypočteného  $z$  skóre. To se provádí pomocí následujícího vztahu:

$$PD = p(B|x_i) = \frac{1}{1 + \frac{1 - \pi_B}{\pi_B} e^{z_i - \alpha}}, \quad (3.30)$$

kde  $z_i$  je veličina definovaná vztahem (3.17),  $\pi_B$  představuje předchozí pravděpodobnost úpadku (ta je často zjišťována na základě historických zkušeností banky se subjekty s obdobnou charakteristikou) a  $\alpha = \frac{1}{2} \gamma'(x_A - x_B)$ , kde  $x_A$  a  $x_B$  jsou vektory středních hodnot nezávislých proměnných pro skupiny zdravých dlužníků a dlužníků v úpadku.

### 3.3.4 Výběr diskriminačních proměnných

Výběr diskriminačních proměnných může být obecně proveden pomocí dvou metod:

- a) přímá metoda,
- b) postupná metoda.

V případě přímé metody jsou diskriminační proměnné vybrány na základě teoretického zdůvodnění a příslušné parametry jsou odhadnuty společně. V druhém případě máme seznam více nezávislých proměnných (tzv. kandidátů), ze kterého proměnné vybíráme na bázi diskriminační výkonnosti. Tento výběr pak může být proveden dvěma způsoby, a to zahrnutím všech diskriminačních proměnných do modelu a následným vyloučením těch s nejnižší diskriminační silou (*backward elimination*), anebo zahrnutím jen jediné proměnné do modelu a následným postupným přidáváním těch, které diskriminační sílu modelu nejvíce zlepšují (*forward selection*). Rovněž existuje i kombinace obou metod.

Další významnou metodou, jež se dá použít při výběru diskriminačních proměnných, je tzv. **korelační analýza**. Cílem této analýzy je zjistit, jak jsou mezi sebou jednotlivé diskriminační proměnné statisticky závislé. K tomu nám stačí sestavit prostou korelační matici. Je důležité sledovat to, aby do modelu nebyly zahrnuty takové proměnné, které jsou mezi sebou silně zkorelovány, protože ačkoli mohou individuálně vykazovat dobrou výkonnost, jejich společné zahrnutí do modelu může vést k nepřesným výsledkům. Jako kritérium se často uvádí 80-ti % hranice. Pokud je tedy korelace mezi dvěma proměnnými vyšší než 80%, neměly by být obě tyto proměnné do modelu zahrnuty.

### **3.4 Ukazatele použité při sestavování skóringových modelů v aplikační části**

V této části práce budou definovány ukazatele, jež budou následně použity v aplikační části při sestavování jednotlivých skóringových modelů. Jelikož v aplikační části bude pracováno s bankami, budou zde definovány takové poměrové ukazatele, které se používají při finanční analýze bank. Následující ukazatele byly vybrány na základě článku autorů Karminsky a Peresetsky (2007).

### 3.4.1 Ukazatele rentability

Prvním z ukazatelů v této skupině je ukazatel **YAEA**, jež vyjadřuje podíl úrokových výnosů (**II** - *interest income*) na průměrných úročených aktivech (**AIEA** - *average interest earnings assets*), což jsou vlastně různé formy půjček:

$$YAEA = \frac{II}{AIEA}. \quad (3.31)$$

Dalším ukazatelem je ukazatel **CIBL**, který poměří nákladové úroky (**IE** – *interest expenses*) s průměrnými úročenými pasivy (**AIBL** – *average interest bearing liabilities*), což jsou většinou různé druhy vkladů klientů:

$$CIBL = \frac{IE}{AIBL}. \quad (3.32)$$

Ukazatel čisté úrokové marže (**NIM** – *net interest margin*) vyjadřuje podíl rozdílu mezi úrokovými výnosy a úrokovými náklady na průměrných úročených aktivech. Je to hojně užívaný ukazatel, z jehož hodnoty se dají vyčíst informace o obchodní strategii banky, závislost banky na úrokových příjmech a míra angažovanosti banky v úvěrově-depozitních obchodech:

$$NIM = \frac{II - IE}{AIEA}. \quad (3.33)$$

K nejčastěji používaným ukazatelům popisujících ziskovost banky patří rentabilita průměrného kapitálu (**ROAE** – *return on average equity*) a rentabilita průměrných aktiv (**ROAA** – *return on average assets*). Průměrný vlastní kapitál či průměrná aktiva nejsnadněji získáme jako aritmetický průměr jejich hodnot na počátku a na konci sledovaného období.

Zatímco ukazatel ROAE představuje míru výnosu, jež patří akcionářům banky, ukazatel ROAA popisuje schopnost banky využívat aktiva k tvorbě zisku, a proto zajímá hlavně řídící management banky:

$$ROAE = \frac{\text{net income}}{\text{average shareholder's equity}}, \quad (3.34)$$

$$ROAA = \frac{\text{net income}}{\text{average assets}}. \quad (3.35)$$

Posledním ukazatelem této skupiny, který bude použit v aplikační části práce, je ukazatel **IE II**, jež poměří nákladové úroky s nákladovými výnosy:

$$IE II = \frac{IE}{II}. \quad (3.36)$$

### 3.4.2 Ukazatele výkonnosti

Hlavním ukazatelem ze skupiny ukazatelů výkonnosti je ukazatel **CIR** (*cost to income ratio*), jež se vypočítá jako podíl provozních nákladů (**OE** – *operating expenses*) a provozních výnosů (**OI** – *operating income*):

$$CIR = \frac{OE}{OI}. \quad (3.37)$$

Dalším ukazatelem je ukazatel **PE OI**, jež vyjadřuje, jakou část z provozních výnosů zaujímají osobní náklady (**PE** – *personnel expenses*):

$$PE OI = \frac{PE}{OI}. \quad (3.38)$$

### 3.4.3 Ukazatele kvality aktiv

Nejdůležitějším ukazatelem této skupiny je ukazatel **PL GL**, jež poměří problémové půjčky (**PL** – *problem loans*) s celkovými půjčkami (**GL** – *gross loans*). Za problémové půjčky můžeme označit takové půjčky, které jsou více než 90 dní po splatnosti (*accruing loans past due 90 days or more*), nevýdělečné půjčky (*nonaccrual loans*) nebo půjčky se zhoršeným ratingem (*impaired loans*):

$$PL\ GL = \frac{PL}{GL}. \quad (3.39)$$

Ukazatel **LLR GL** se vypočte jako podíl rezerv na ztrátové půjčky (**LLR** – *loan loss reserve*) a celkových půjček:

$$LLR\ GL = \frac{LLR}{GL}. \quad (3.40)$$

Jako poslední ukazatel této skupiny bude definován ukazatel **PL EQ LLR**. Tento ukazatel poměruje problémové půjčky se součtem vlastního kapitálu (**EQ** – *shareholder's equity*) a rezerv na ztrátové půjčky:

$$PL\ EQ\ LLR = \frac{PL}{EQ + LLR}. \quad (3.41)$$

#### 3.4.4 Ukazatele kapitálové přiměřenosti

Prvním ukazatelem řadícím se do skupiny ukazatelů kapitálové přiměřenosti je ukazatel **EQ TA**, jež poměruje vlastní kapitál s celkovou hodnotou aktiv (**TA** – *total assets*). Tento ukazatel nepočítá s rozdílnou rizikovostí jednotlivých aktiv banky:

$$EQ\ TA = \frac{EQ}{TA}. \quad (3.42)$$

Dalším ukazatelem je ukazatel **CAR**, jež se vypočte jako podíl celkového kapitálu banky (**TC** – *total capital*) a rizikově vážených aktiv (**RWA** – *risk weighted assets*). Tento ukazatel by měl nabývat hodnot větších než 8% :

$$CAR = \frac{TC}{RWA}. \quad (3.43)$$

Ukazatel ***T1*** (*tier 1 ratio*) je obdoba ukazatele CAR s tím rozdílem, že se s rizikově váženými aktivy nepoměřuje celkový kapitál banky ale pouze kapitál tier 1 (**T1 C** – *tier 1, core capital*). Do kapitálu tier 1 se řadí základní kapitál banky, emisní ážio, povinné rezervní fondy, ostatní rezervní fondy ze zisku, nerozdělený zisk z předchozích období po zdanění a zisk běžného období snížený o předpokládané dividendy a snížený o odečitatelné položky:

$$T1 = \frac{T1\ C}{RWA}. \quad (3.44)$$

Jako poslední ukazatel definujeme ukazatel ***D EQ***, jež poměřuje celkové vklady (**D** – *total deposits*) s vlastním kapitálem:

$$D\ EQ = \frac{D}{EQ}. \quad (3.45)$$

## 4 Aplikace na vybrané banky

V této kapitole budou představená teoreticko-metodologická východiska aplikována na vzorek bank. Nejdříve budou popsána vstupní data, která jsou získána z finančních výkazů jednotlivých bank. Dále budou odhadnuty tři skóringové modely – logit model, probit model a model sestavený na základě lineární diskriminační analýzy (LDA model). Statistická významnost těchto modelů a jednotlivých parametrů bude testována vhodnými testy. K odhadu a testování logit a probit modelu bude použit statistický software *STATA<sup>tm</sup> 10.0 MP*.

Tyto tři sestavené skóringové modely budou následně aplikovány na kontrolní vzorek bank s cílem určit nejvhodnější model pro predikci úpadku.

### 4.1 Vstupní data

Součástí této podkapitoly je charakteristika vzorku bank, na který bude představená metodologie aplikována, dále popis vybraných finančních ukazatelů a taktéž korelační analýza ukazatelů.

#### 4.1.1 Popis bank obsažených ve vzorku

Při sestavování jednotlivých modelů je pracováno se vzorkem, který obsahuje celkem 298 amerických komerčních bank. Seznam těchto bank je uveden v **příloze č. 1**. Tyto banky jsou rozděleny do dvou skupin: **skupina A** (137 *non-default* bank) a **skupina B** (161 *default* bank).

V této práci jsou default bankami myšleny takové finanční instituce, které vstoupily do insolvenčního řízení nebo byly účastny jiných finančně-restrukturalizačních procesů (např. byly převzaty jinou bankou anebo byly zestátněny). Vzorek bank pro obě skupiny byl vybrán náhodně na základě veřejně dostupných informací.

## 4.1.2 Charakteristika finančních ukazatelů

Pro sestavení modelů je použito celkem 16 finančních ukazatelů, které popisují finanční zdraví jednotlivých bank. Jsou to ukazatele popisující velikost banky, rentabilitu, výkonnost, kvalitu jejích aktiv a také kapitálovou přiměřenost. Přehled použitých ukazatelů a jejich rozdělení do základních oblastí udává následující tabulka (Tab. 4.1).

Tab. 4.1 Přehled použitých ukazatelů a jejich zařazení do základních oblastí

Zkratka	Ukazatel	Oblast
LTA	Logaritmus celkových aktiv	Velikost
YAEA	Výnosové úroky / Průměrná úročená aktiva (%)	Rentabilita
CIBL	Nákladové úroky / Průměrná úročená pasiva (%)	Rentabilita
NIM	(Výnosové úroky – Nákladové úroky) / Průměrná úročená aktiva (%)	Rentabilita
ROAA	Rentabilita průměrných aktiv (%)	Rentabilita
ROAE	Rentabilita průměrného kapitálu (%)	Rentabilita
IE II	Nákladové úroky / Výnosové úroky (%)	Rentabilita
CIR	Provozní náklady / Provozní výnosy (%)	Výkonnost
PE OI	Osobní náklady / Provozní výnosy (%)	Výkonnost
PL GL	Problémové půjčky / Celkové půjčky (%)	Kvalita aktiv
LLR GL	Rezervy na ztrátové půjčky / Celkové půjčky (%)	Kvalita aktiv
PL EQ LLR	Problémové půjčky / (Vlastní kapitál + Rezervy na ztrátové půjčky) (%)	Kvalita aktiv
T1	Tier 1 kapitál / Rizikově vážená aktiva (%)	Kapitálová přiměřenost
EQ TA	Vlastní kapitál / Celková aktiva (%)	Kapitálová přiměřenost
CAR	Celkový kapitál / Rizikově vážená aktiva (%)	Kapitálová přiměřenost
D EQ	Celkové vklady / Vlastní kapitál	Kapitálová přiměřenost

Jednotlivé vypočtené finanční ukazatele pro všechny banky jsou součástí **přílohy č. 2**. Střední hodnoty vybraných finančních ukazatelů pro obě skupiny bank jsou uvedeny v následující tabulce (Tab. 4.2).

Tab. 4.2 Střední hodnota finančních ukazatelů pro non-default a default banky

Non-default banky				Default banky			
fin. ukazatel	stř. hodnota	fin. ukazatel	stř. hodnota	fin. ukazatel	stř. hodnota	fin. ukazatel	stř. hodnota
$x_1$ : LTA	15,80	$x_9$ : PE OI	23,21%	$x_1$ : LTA	11,91	$x_9$ : PE OI	26,62%
$x_2$ : YAEA	5,81%	$x_{10}$ : PL GL	3,71%	$x_2$ : YAEA	6,65%	$x_{10}$ : PL GL	15,15%
$x_3$ : CIBL	3,30%	$x_{11}$ : LLR GL	1,96%	$x_3$ : CIBL	3,69%	$x_{11}$ : LLR GL	3,24%
$x_4$ : NIM	3,56%	$x_{12}$ : PL EQ LLR	27,42%	$x_4$ : NIM	3,15%	$x_{12}$ : PL EQ LLR	39,09%
$x_5$ : ROAA	1,14%	$x_{13}$ : T1	10,70%	$x_5$ : ROAA	-4,31%	$x_{13}$ : T1	7,84%
$x_6$ : ROAE	7,62%	$x_{14}$ : EQ TA	10,92%	$x_6$ : ROAE	-47,53%	$x_{14}$ : EQ TA	5,58%
$x_7$ : IE II	37,87%	$x_{15}$ : CAR	12,60%	$x_7$ : IE II	55,43%	$x_{15}$ : CAR	8,35%
$x_8$ : CIR	86,81%	$x_{16}$ : D EQ	7,86	$x_8$ : CIR	115,54%	$x_{16}$ : D EQ	15,32



### 4.1.3 Korelační analýza ukazatelů

Korelační matice všech šestnácti finančních ukazatelů je uvedena v **příloze č. 3**. Z výsledků je patrné, že žádná dvojice ukazatelů mezi sebou není zkorelována natolik, aby hodnota korelace překročila hranici 80%. Z hlediska korelační analýzy tak do modelů mohou být zahrnuty všechny ukazatele.

Nejvyšší hodnota korelace je necelých 75% mezi ukazateli *ROAA* a *ROAE*, ale v další části práce se přesvědčíme, že ani tyto dva ukazatele nejsou společně zahrnuty do žádného sestaveného modelu.

## 4.2 Logit model

V této podkapitole bude nejprve odhadnut logit model, dále budou testovány jednotlivé parametry modelu pomocí likelihood-ratio testu a Waldova testu a nakonec bude součástí také grafické znázornění pravděpodobností úpadků jednotlivých bank.

### 4.2.1 Odhad logit modelu

Odhadnutý logit model je znázorněn v následující tabulce (Tab. 4.3), jež je výstupem z již zmíněného statistického softwaru *STATA*.

Tab. 4.3 Odhadnutý logit model

begin with empty model						
p = 0.0000	<	0.0500	adding	x10_PL_GL		
p = 0.0000	<	0.0500	adding	x1_LTA		
p = 0.0170	<	0.0500	adding	x16_D_EQ		
p = 0.0181	<	0.0500	adding	x12_PL_EQ_LLR		
p = 0.0279	<	0.0500	adding	x6_ROAE		
p = 0.8739	>=	0.1000	removing	x16_D_EQ		
p = 0.5968	>=	0.1000	removing	x12_PL_EQ_LLR		
Logistic regression				Number of obs	=	298
				LR chi2(3)	=	395.95
				Prob > chi2	=	0.0000
Log likelihood = -7.6143581				Pseudo R2	=	0.9630
Status	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x10_PL_GL	87.69232	35.89412	2.44	0.015	17.34115	158.0435
x1_LTA	-4.477891	1.786542	-2.51	0.012	-7.979449	-.9763325
x6_ROAE	-36.10652	15.96435	-2.26	0.024	-67.39607	-4.816973
_cons	52.08043	21.22815	2.45	0.014	10.47401	93.68685

Zdroj: STATA<sup>tm</sup> 10.0 MP

Model tak dle vztahu (3.5) můžeme zapsat jako:

$$\hat{\pi}(x_i) = PD_i = \frac{1}{1 + e^{-(52,08 - 4,48x_{1,i} - 36,11x_{6,i} + 87,69x_{10,i})}},$$

kde  $x_1$  je ukazatel LTA,  $x_6$  ROAE a  $x_{10}$  PL GL.

Při odhadování tohoto modelu byla použita stepwise metoda (*forward selection* následovaná *backward elimination*) s hladinami významnosti  $p_E = 0,05$  pro přidání proměnné do modelu a  $p_R = 0,1$  pro vyloučení proměnné z modelu. Ze sloupce označeného  $P > |z|$  tedy můžeme vyčíst, že všechny čtyři parametry modelu jsou statisticky významné na 5% hladině významnosti.

Logaritmus věrohodnostní funkce (*log-likelihood*) tohoto modelu zjištěný dle vztahu (3.9) má hodnotu -7,61.

## Pseudo $R^2$

Hodnota *pseudo  $R^2$*  činí 0,963. Tato hodnota se sice nedá interpretovat stejně jako koeficient determinace  $R^2$  v případě lineární regrese, neboť odhady maximální věrohodnosti (*maximum likelihood estimates*) nejsou vypočteny na základě minimalizace rozptylu, ale i zde platí, že vyšší hodnota znamená lepší vypovídací schopnost modelu. *Pseudo  $R^2$*  (někdy také označováno jako *McFadden's pseudo  $R^2$* ) se vypočte dle následujícího vztahu:

$$pseudo R^2 = 1 - \frac{L_1}{L_0}, \quad (4.1)$$

kde  $L_1$  značí log-likelihood hodnotu „plného“ modelu a  $L_0$  modelu obsahujícího pouze konstantu (tzv. *intercept model*). V našem případě činí hodnota  $L_0$  -205,59. Potom tedy dle vztahu (4.1)

$$pseudo R^2 = 1 - \frac{(-7,61)}{(-205,59)} = 0,963.$$

## Konfidenční interval

Konfidenční intervaly pro jednotlivé parametry jsou rovněž uvedeny v Tab. 4.3, kde byla zvolena 5% hladina významnosti  $\alpha$ . Například pro ukazatel PL GL bychom tento interval ručně spočetli dle vztahu (3.15) následovně:

$$87,69 \pm 1,96 \cdot 35,89 = (17,34; 158,04).$$

### 4.2.2 Testování významnosti parametrů

Testování odhadnutých parametrů logit modelu bude provedeno pomocí likelihood-ratio testu a Waldova testu.

#### Likelihood-ratio test

Tento test je součástí výstupu logistické regrese softwaru *STATA* a můžeme ho tedy vyčíst z řádků „*LR chi2(3)*“ a „*Prob > chi2*“ v Tab. 4.3. Napřed si stanovme nulovou a alternativní hypotézu:

$$H_0: \hat{\beta}_j = 0 \quad \text{kde } j = 1, 2, 3;$$

$$H_A: \hat{\beta}_1 \neq 0 \vee \hat{\beta}_2 \neq 0 \vee \hat{\beta}_3 \neq 0 \quad (\text{alespoň jeden z parametrů je simultánně nenulový}).$$

Statistiku LR vypočteme dle vztahu (3.12) následovně:

$$LR = -2[(-205,59) - (-7,61)] = 395,95$$

a potom

$$P[\chi^2(3) > 395,95] = 0,0000.$$

Vzhledem k tomu, že  $0,0000 < 0,05$ , zamítáme nulovou hypotézu a přijímáme hypotézu alternativní, tedy že alespoň jeden z parametrů modelu je různý od nuly a model tak je statisticky významný na 5% hladině významnosti.

## Waldův test

Nulová a alternativní hypotéza je u tohoto testu totožná s předchozím testem, tedy:

$$H_0: \hat{\beta}_j = 0 \quad \text{kde } j = 1, 2, 3;$$

$$H_A: \hat{\beta}_1 \neq 0 \vee \hat{\beta}_2 \neq 0 \vee \hat{\beta}_3 \neq 0 \quad (\text{alespoň jeden z parametrů je simultánně nenulový}).$$

Hodnoty Waldova testu jsou v případě, kdy testujeme všechny parametry současně, uvedeny v následující tabulce (Tab. 4.4).

*Tab. 4.4 Waldův test logit modelu*

```
( 1)  x10_PL_GL = 0
( 2)  x1_LTA = 0
( 3)  x6_ROAE = 0

      chi2( 3) =      6.99
    Prob > chi2 =    0.0421
```

Zdroj: STATA<sup>tm</sup> 10.0 MP

Hodnota 0,0421 je nižší než 0,05, zamítáme tedy nulovou hypotézu a přijímáme hypotézu alternativní, tedy že alespoň jeden z parametrů modelu je různý od nuly a model tak je statisticky významný na 5% hladině významnosti.

### 4.2.3 Grafické znázornění PD jednotlivých bank

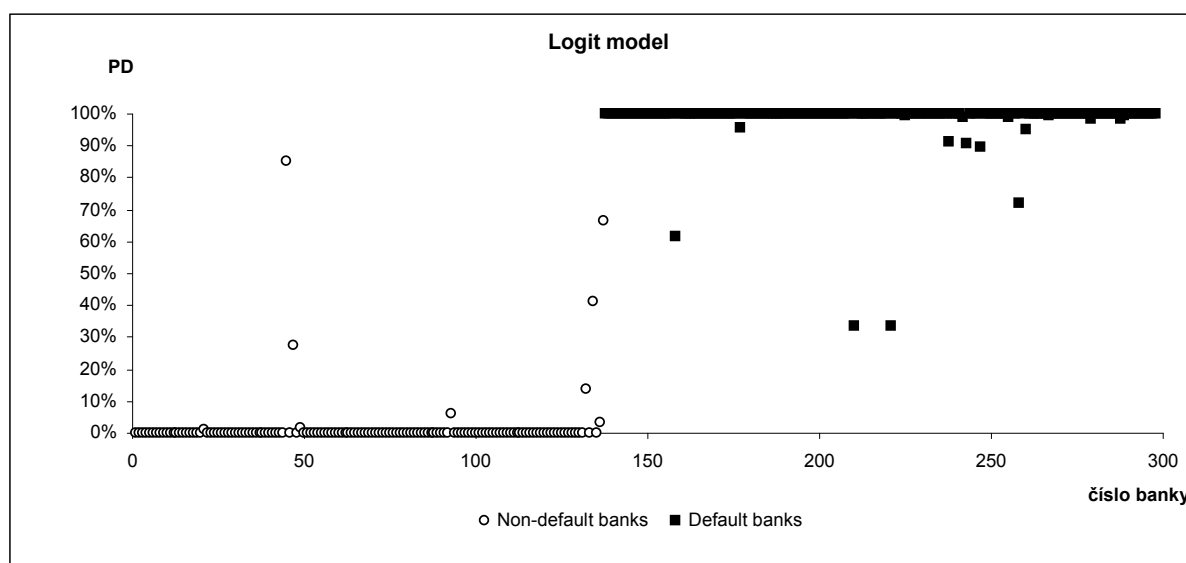
V následující tabulce (Tab. 4.5) jsou uvedeny odhadnuté střední hodnoty pravděpodobnosti úpadku pro obě skupiny bank.

*Tab. 4.5 Odhadnuté střední hodnoty PD pro logit model*

Non-default banky	střední hodnota	Default banky	střední hodnota
<b>PD</b>	1,80%	<b>PD</b>	98,47%

Graf 4.1 zobrazuje pravděpodobnost úpadku každé banky ze vzorku, kde na ose x je pořadové číslo banky a na ose y je pravděpodobnost úpadku.

Graf 4.1 Odhadnuté PD jednotlivých bank logit modelu



### 4.3 Probit model

Součástí této podkapitoly bude odhad probit modelu, testování jednotlivých parametrů modelu a znovu také grafické znázornění pravděpodobností úpadků jednotlivých bank sestavených na základě tohoto modelu.

#### 4.3.1 Odhad probit modelu

Odhadnutý probit model je znázorněn v následující tabulce (Tab. 4.6).

Tab. 4.6 Odhadnutý probit model

```
begin with empty model
p = 0.0000 < 0.0500 adding x1_LTA
p = 0.0000 < 0.0500 adding x5_ROAA
p = 0.0001 < 0.0500 adding x9_PE_OI
p = 0.0003 < 0.0500 adding x6_ROAE
p = 0.3117 >= 0.1000 removing x5_ROAA
p = 0.0273 < 0.0500 adding x14_EQ_TA
```

Probit regression

Number of obs = 298  
LR chi2(4) = 394.11  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.9585

Log likelihood = -8.5362047

Status	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1_LTA	-1.702366	.4842775	-3.52	0.000	-2.651532	-.7531992
x6_ROAE	-18.26809	6.159848	-2.97	0.003	-30.34117	-6.195008
x9_PE_OI	-17.04526	6.168434	-2.76	0.006	-29.13517	-4.955347
x14_EQ_TA	-82.14669	37.22711	-2.21	0.027	-155.1105	-9.182897
_cons	33.75012	9.742617	3.46	0.001	14.65494	52.8453

Zdroj: STATA<sup>tm</sup> 10.0 MP

Tento model tak podle vztahu (3.7) můžeme zapsat ve tvaru:

$$\hat{y}_i = PD_i = \Phi(33,75 - 1,7x_{1,i} - 18,27x_{6,i} - 17,05x_{9,i} - 82,15x_{14,i}),$$

kde  $x_1$  je ukazatel LTA,  $x_6$  ROAE,  $x_9$  PE OI a  $x_{14}$  EQ TA.

Při odhadování tohoto modelu byla rovněž použita stepwise metoda (*forward selection* následovaná *backward elimination*) se stejnými hladinami významnosti jako u logit modelu, tedy  $p_E = 0,05$  pro přidání proměnné do modelu a  $p_R = 0,1$  pro vyloučení proměnné z modelu. Opět ze sloupce označeného jako  $P > |z|$  můžeme vyčíst, že všech pět parametrů modelu je statisticky významných na 5% hladině významnosti.

Logaritmus věrohodnostní funkce (*log-likelihood*) tohoto modelu zjištěný dle vztahu (3.9) má hodnotu -8,54.

## Pseudo $R^2$

Hodnota *pseudo  $R^2$*  činí 0,9585. Log-likelihood modelu obsahujícího pouze konstantu  $L_0$  je pochopitelně stejný jako u logit modelu, jeho hodnota tedy činí -205,59. Dle vztahu (4.1) tedy hodnotu *pseudo  $R^2$*  probit modelu můžeme vypočítat následovně:

$$pseudo\ R^2 = 1 - \frac{(-8,54)}{(-205,59)} = 0,9585.$$

## Konfidenční interval

Pro konfidenční intervaly jednotlivých parametrů probit modelu byla opět zvolena 5% hladina významnosti  $\alpha$ . Například pro ukazatel ROAE bychom tento interval ručně spočetli dle vztahu (3.15) následovně:

$$-18,27 \pm 1,96 \cdot 6,16 = (-30,34; -6,2).$$

### 4.3.2 Testování významnosti parametrů

Testování odhadnutých parametrů probit modelu bude provedeno pomocí likelihood-ratio testu a Waldova testu.

#### Likelihood-ratio test

Tento test je opět součástí výstupu softwaru *STATA*, tentokrát v případě probit regrese, a můžeme ho tedy znovu vyčíst v Tab. 4.6. Nulová a alternativní hypotéza jsou definovány stejně jako u logit modelu s ohledem na to, že probit model obsahuje již čtyři ukazatele:

$$H_0: \hat{\beta}_j = 0 \quad \text{kde } j = 1, 2, 3, 4;$$

$$H_A: \hat{\beta}_1 \neq 0 \vee \dots \vee \hat{\beta}_4 \neq 0 \quad (\text{alespoň jeden z parametrů je simultánně nenulový}).$$

V případě probit modelu statistiku LR vypočteme dle vztahu (3.12) následovně:

$$LR = -2[(-205,59) - (-8,54)] = 394,11$$

a potom

$$P[\chi^2(4) > 394,11] = 0,0000.$$

Vzhledem k tomu, že  $0,0000 < 0,05$ , zamítáme nulovou hypotézu a přijímáme hypotézu alternativní, tedy že alespoň jeden z parametrů modelu je různý od nuly a model tak je statisticky významný na 5% hladině významnosti.

#### Waldův test

Nulová a alternativní hypotéza je u tohoto testu znovu definována stejným způsobem jako u předcházejícího testu, tedy:

$$H_0: \hat{\beta}_j = 0 \quad \text{kde } j = 1, 2, 3, 4;$$

$$H_A: \hat{\beta}_1 \neq 0 \vee \dots \vee \hat{\beta}_4 \neq 0 \quad (\text{alespoň jeden z parametrů je simultánně nenulový}).$$

Hodnoty Waldova testu probit modelu jsou v případě, kdy opět testujeme všechny parametry současně, uvedeny v následující tabulce (Tab. 4.7).

Tab. 4.7 Waldův test probit modelu

```
( 1) x1_LTA = 0
( 2) x6_ROAE = 0
( 3) x9_PE_OI = 0
( 4) x14_EQ_TA = 0

      chi2( 4) =    12.88
      Prob > chi2 =    0.0119
```

Zdroj: STATA<sup>tm</sup> 10.0 MP

Hodnota 0,0119 je nižší než 0,05, zamítáme tedy nulovou hypotézu a přijímáme hypotézu alternativní, tedy že alespoň jeden z parametrů modelu je různý od nuly a model tak je statisticky významný na 5% hladině významnosti.

### 4.3.3 Grafické znázornění PD jednotlivých bank

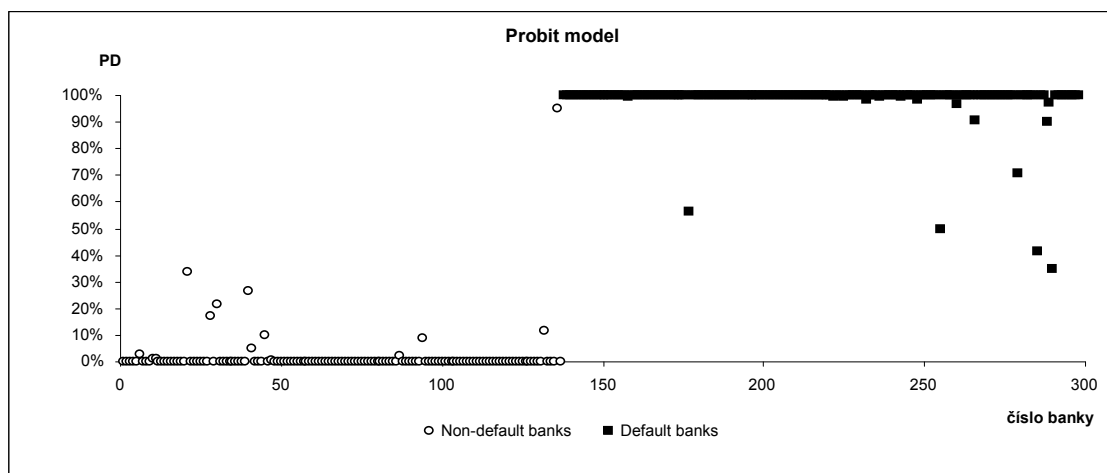
V následující tabulce (Tab. 4.8) jsou uvedeny odhadnuté střední hodnoty pravděpodobnosti úpadku pro probit model pro obě skupiny bank.

Tab. 4.8 Odhadnuté střední hodnoty PD pro probit model

Non-default banky	střední hodnota	Default banky	střední hodnota
<b>PD</b>	1,74%	<b>PD</b>	98,27%

Graf 4.2 znovu zobrazuje pravděpodobnost úpadku každé banky ze vzorku, tentokrát v rámci probit modelu.

Graf 4.2 Odhadnuté PD jednotlivých bank probit modelu





## 4.4 Lineární diskriminační analýza

V této podkapitole bude nejprve odhadnut LDA model, pomocí  $t$ -testů a  $F$ -testu bude zkoumána statistická významnost jednotlivých parametrů modelu a významnost modelu jako celku. Dále bude propočtena míra úspěšnosti tohoto modelu (Wilksovo lambda) a součástí této podkapitoly bude rovněž grafické znázornění pravděpodobností úpadků jednotlivých bank sestavených na základě tohoto modelu.

### 4.4.1 Odhad LDA modelu

Hodnoty vektoru  $\gamma$  predikčního modelu sestaveného na základě lineární diskriminační analýzy jsou vypočteny dle vztahu (3.19). Výběr diskriminačních proměnných v modelu byl proveden pomocí postupné metody (*backward elimination*) a následně tak byl získán vektor  $\gamma'$ .

$$\gamma = \begin{bmatrix} 1,03 \\ -22,17 \\ 35,52 \\ -95,15 \\ 41,86 \\ -0,01 \\ -15,68 \\ 1,79 \\ 0,68 \\ -8,84 \\ -0,19 \\ -0,06 \\ -32,95 \\ 72,27 \\ 6,01 \\ 0,01 \end{bmatrix} ; \quad \gamma' = \begin{bmatrix} 1,14 \\ 0,00 \\ 0,00 \\ 0,00 \\ 24,91 \\ 0,00 \\ -5,76 \\ 0,00 \\ 0,00 \\ -9,26 \\ 0,00 \\ 0,00 \\ 0,00 \\ 52,57 \\ 0,00 \\ 0,00 \end{bmatrix}$$

Výsledný vektor  $\gamma'$  již obsahuje pouze pět ukazatelů. Obecný LDA model sestavený dle vztahu (3.17) poté vypadá následovně:

$$\hat{z}_i = 1,14x_{1,i} + 24,91x_{5,i} - 5,76x_{7,i} - 9,26x_{10,i} + 52,57x_{14,i},$$

kde  $x_1$  je ukazatel LTA,  $x_5$  ROAA,  $x_7$  IE II,  $x_{10}$  PL GL a  $x_{14}$  EQ TA.

#### 4.4.2 Statistická verifikace odhadnutých parametrů a modelu

Statistická významnost jednotlivých odhadnutých parametrů LDA modelu bude testována pomocí  $t$ -testů, statistická významnost modelu jako celku pak pomocí  $F$ -testu.

##### **t-test**

Nulová a alternativní hypotéza  $t$ -testu jsou v našem případě definovány následujícím způsobem:

$$H_0: \hat{\beta}_i = 0 ;$$

$$H_A: \hat{\beta}_i \neq 0 \quad \text{kde } i = 1, 2, 3, 4, 5 .$$

Statistiky  $t$ -testu pro všechny parametry modelu jsou uvedeny v následující tabulce (Tab. 4.9).

Tab. 4.9  $t$ -test LDA modelu

	<b>beta</b>	<b>SEbeta</b>	<b><math>t^{vyp}</math></b>	<b>hodnota P</b>
<b><math>x_1</math>: LTA</b>	1,14	0,3306	3,4487	0,0006
<b><math>x_5</math>: ROAA</b>	24,91	11,4084	2,1835	0,0298
<b><math>x_7</math>: IE II</b>	-5,76	2,8446	-2,0249	0,0438
<b><math>x_{10}</math>: PL GL</b>	-9,26	3,8457	-2,4079	0,0167
<b><math>x_{14}</math>: EQ TA</b>	52,57	22,2995	2,3575	0,0191

Pro tento test byla zvolena 5% hladina významnosti  $\alpha$ . Kritická hodnota ( $t^{krit}$ ) je vypočtena pomocí funkce **TINV** v programu *MS Excel* na 5% hladině významnosti a tato hodnota v našem případě činí 1,9681.

Z tabulky 4.9 lze vyčíst, že hodnoty  $t^{vyp}$  (brány v absolutní hodnotě) u všech parametrů jsou vyšší než vypočtená kritická hodnota  $t^{krit}$ . Dále v této tabulce vidíme, že všechny „hodnoty P“ jsou nižší než zvolená hladina významnosti. Díky těmto skutečnostem

zamítáme nulovou hypotézu a přijímáme hypotézu alternativní – všechny propočtené parametry leží v kritické oblasti a jsou tedy statisticky významné na 5% hladině významnosti.

## F-test

U tohoto testu je v našem případě nulová a alternativní hypotéza definována podobně jako u likelihood-ratio testu a Waldova testu, tedy:

$$H_0: \hat{\beta}_i = 0 \quad \text{kde } i = 1, 2, 3, 4, 5;$$

$$H_A: \hat{\beta}_1 \neq 0 \vee \dots \vee \hat{\beta}_5 \neq 0 \quad (\text{alespoň jeden z parametrů je simultánně nenulový}).$$

Statistika  $F$ -testu LDA modelu je uvedena v následující tabulce (Tab. 4.10).

Tab. 4.10  $F$ -test LDA modelu

<b>Regresní statistika</b>					
<b>R</b>	0,885653				
<b>R<sup>2</sup></b>	0,784381				
<b>SE</b>	17,117242				
<b>T</b>	298				

<b>ANOVA</b>					
	<b>df</b>	<b>SS</b>	<b>MS</b>	<b>F<sup>vyp</sup></b>	<b>významnost F</b>
<b>regrese</b>	5	58,057407	11,611481	227,699869	1,13581E-98
<b>rezidua</b>	293	14,941440	0,050994		
<b>celkem</b>	298	74,016778			

Znovu byla pro tento test zvolena 5% hladina významnosti  $\alpha$ . Kritická hodnota ( $F^{\text{krit}}$ ) je vypočtena pomocí funkce **FINV** v programu *MS Excel* na 5% hladině významnosti a tato hodnota v našem případě činí 2,2448.

Z tabulky 4.10 můžeme vyčíst, že hodnota  $F^{\text{vyp}}$  je vyšší než vypočtená kritická hodnota  $F^{\text{krit}}$ . Dále v této tabulce vidíme, že „hodnota P“ (v tabulce označena jako „významnost F“) je nižší než zvolená hladina významnosti. Díky těmto skutečnostem opět zamítáme nulovou hypotézu a přijímáme hypotézu alternativní – odhadnutý model jako celek je statisticky významný na 5% hladině významnosti.

Dále je z této tabulky možno vyčíst, že koeficient determinace  $R^2$ , což je v podstatě míra úspěšnosti modelu (viz 4.4.3), činí 0,7844.

#### 4.4.3 Míra úspěšnosti modelu

Hodnota Wilksovy lambdy spočítaná dle vztahu (3.29) činí

$$\Lambda = \frac{796,819 + 1592,515}{11079,401} = 0,2156.$$

Po odečtení této hodnoty od jedničky získáváme skutečnou míru úspěšnosti modelu, tedy  $1 - 0,2156 = 0,7844$ , která se shoduje s koeficientem determinace  $R^2$ . Míra úspěšnosti tohoto modelu tedy činí 78,44 %.

#### 4.4.4 Grafické znázornění PD jednotlivých bank

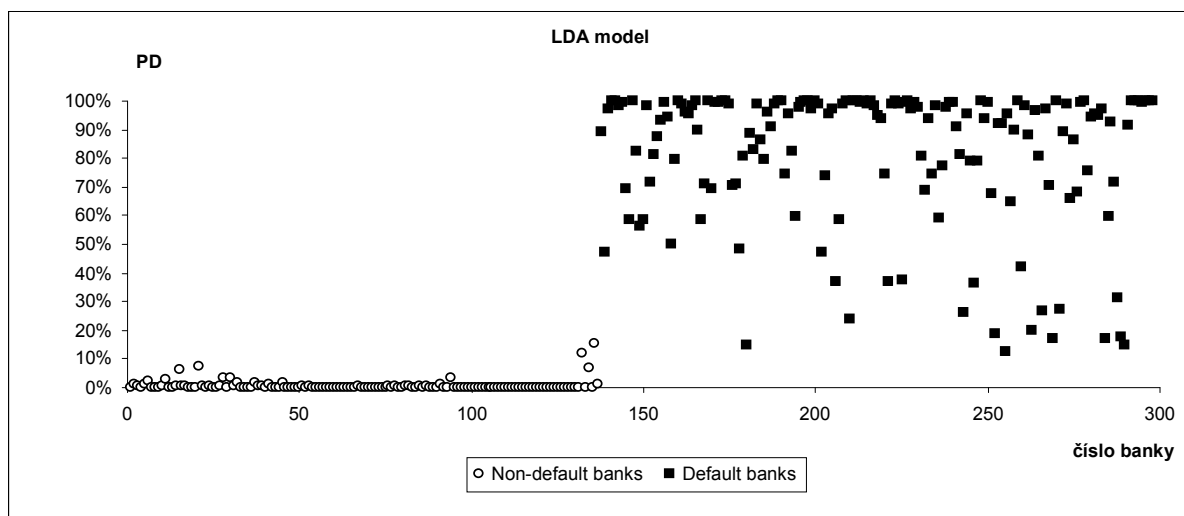
Vypočtené střední hodnoty  $z$  skóre dle vztahu (3.17) a pravděpodobnosti úpadku dle vztahu (3.30) pro obě skupiny bank jsou uvedeny v následující tabulce (Tab. 4.11).

*Tab. 4.11 Odhadnuté střední hodnoty  $z$  skóre a PD pro LDA model*

Non-default banky	střední hodnota	Default banky	střední hodnota
<b><math>z</math></b>	21,498	<b><math>z</math></b>	10,831
<b><math>PD</math></b>	0,67%	<b><math>PD</math></b>	81,65%

Graf 4.3 opět zobrazuje pravděpodobnost úpadku každé banky ze vzorku.

Graf 4.3 Odhadnuté PD jednotlivých bank LDA modelu



## 4.5 Srovnání odhadnutých modelů

Základní rovnice pro jednotlivé odhadnuté modely vypadají následovně:

- pro logit model 
$$\hat{\pi}(x_i) = PD_i = \frac{1}{1 + e^{-(52,08 - 4,48x_{1,i} - 36,11x_{6,i} + 87,69x_{10,i})}},$$
- pro probit model 
$$\hat{y}_i = PD_i = \Phi(33,75 - 1,7x_{1,i} - 18,27x_{6,i} - 17,05x_{9,i} - 82,15x_{14,i}),$$
- a pro LDA model 
$$\hat{z}_i = 1,14x_{1,i} + 24,91x_{5,i} - 5,76x_{7,i} - 9,26x_{10,i} + 52,57x_{14,i}.$$

V následující tabulce (Tab. 4.12) je uveden počet a seznam ukazatelů obsažených v jednotlivých modelech, hodnoty *pseudo*  $R^2$  /  $R^2$  a střední hodnoty pravděpodobnosti úpadku pro obě skupiny bank.

Tab. 4.12 Srovnání jednotlivých odhadnutých modelů

	Logit model		Probit model		LDA model	
<b>počet ukazatelů</b>	3		4		5	
<b>ukazatele</b>	LTA, ROAE, PL GL		LTA, ROAE, PE OI, EQ TA		LTA, ROAA, IE II, PL GL, EQ TA	
<b>pseudo <math>R^2</math> / <math>R^2</math></b>	0,9630		0,9585		0,7844	
<b>E (PD)*</b>	1,80%	98,47%	1,74%	98,27%	0,67%	81,65%

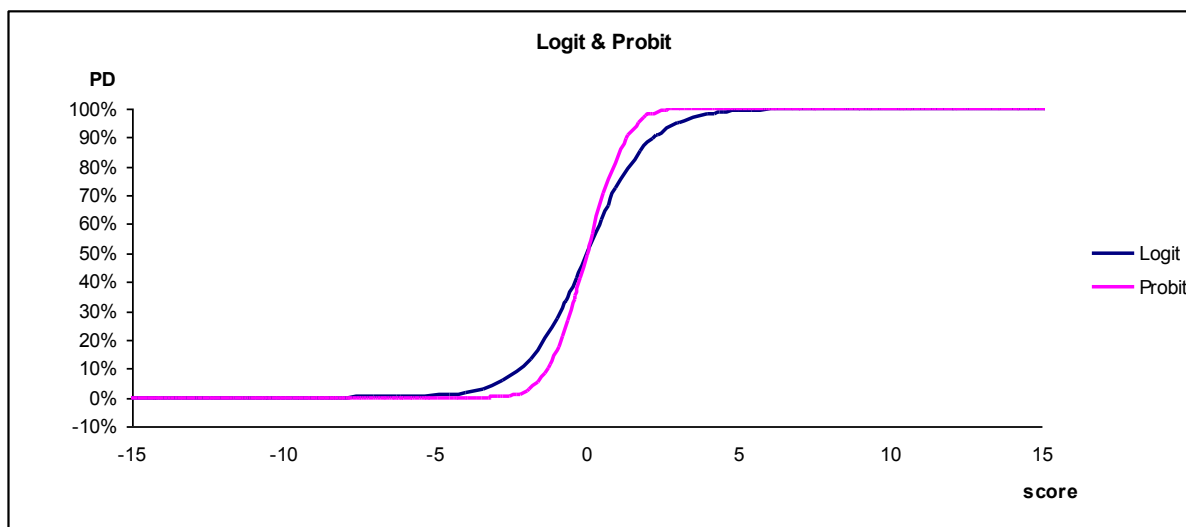
\* První hodnota značí střední hodnotu PD pro non-default banky, druhá pak pro default banky.

Z tabulky můžeme vyčíst, že nejméně ukazatelů (3) obsahuje logit model. Co se týče vypovídací schopnosti logit a probit modelu (*pseudo R<sup>2</sup>*), tak je dosti podobná (96,30% u logit modelu a 95,85% u probit modelu), i když probit model má o jeden ukazatel navíc. Také srovnání středních hodnot pravděpodobnosti úpadku pro obě skupiny bank je u těchto dvou modelů téměř identické. Tyto skutečnosti potvrzují teoretická východiska, že modely logit a probit si jsou velice podobné.

Oproti těmto dvěma modelům obsahuje model odhadnutý pomocí lineární diskriminační analýzy již pět ukazatelů, ale přesto je jeho vypovídací schopnost nižší (78,44%). Z tohoto hlediska jsou tedy první dva modely pro predikci úpadku lepší. LDA model předpovídá úpadek o něco lépe u non-default bank, ale výrazně hůře u default bank (střední hodnota PD je jen 81,65%).

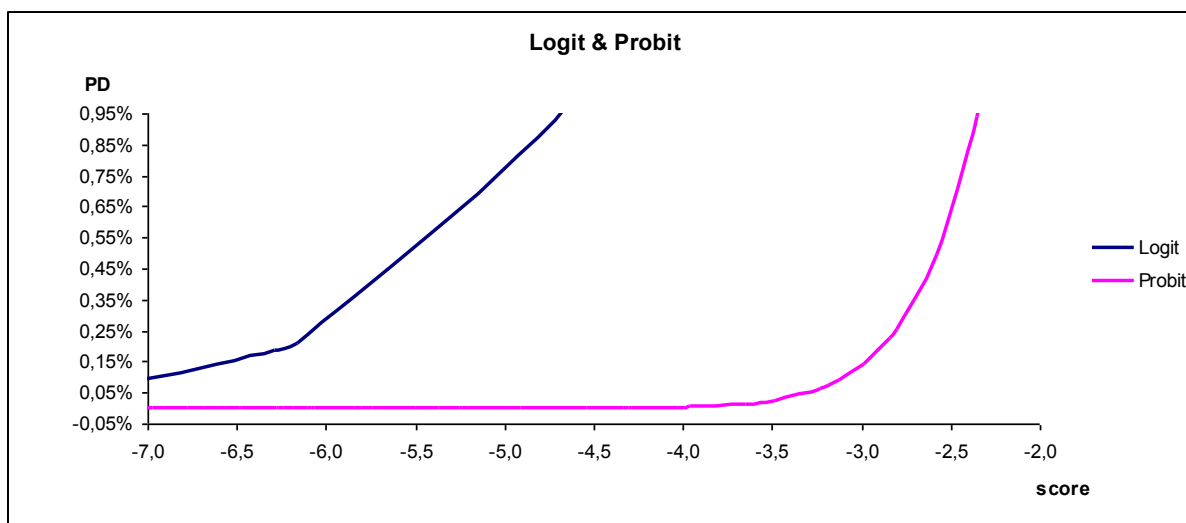
Logit model a probit model jsou graficky srovnány na následujícím grafu (Graf 4.4).

*Graf 4.4 Srovnání logit a probit modelu*



Na tomto grafu vidíme, že křivky závislosti mezi skórem a pravděpodobností úpadku jsou u obou těchto modelů velice podobné. Rozdíl je v tom, že logit model má mírně plošší konce. Detail funkcí logit a probit modelu je uveden v grafu 4.5.

Graf 4.5 Detail funkcí logit a probit modelu



## 4.6 Aplikace odhadnutých modelů na kontrolním vzorku bank

Odhadnuté modely jsou dále aplikovány na kontrolní vzorek bank, které nebyly použity při sestavování těchto modelů. Tento kontrolní vzorek obsahuje 100 amerických komerčních bank, které jsou opět rozděleny do dvou skupin: **skupina AA** (50 *non-default* bank) a **skupina BB** (50 *default* bank). Seznam bank kontrolního vzorku je uveden v **příloze č. 4**.

Pro každou banku z tohoto kontrolního vzorku je spočteno 7 finančních ukazatelů, které jsou obsaženy v sestavených modelech ( $x_1$  LTA,  $x_5$  ROAA,  $x_6$  ROAE,  $x_7$  IE II,  $x_9$  PE OI,  $x_{10}$  PL GL,  $x_{14}$  EQ TA). Hodnoty těchto vybraných finančních ukazatelů pro jednotlivé banky jsou součástí **přílohy č. 5**. Střední hodnoty těchto ukazatelů pro obě skupiny bank v kontrolním vzorku jsou uvedeny v následující tabulce (Tab. 4.13).

Tab. 4.13 Střední hodnoty vybraných finančních ukazatelů kontrolního vzorku bank

Non-default banky		Default banky	
fin. ukazatel	stř. hodnota	fin. ukazatel	stř. hodnota
$x_1$ : LTA	13,45	$x_1$ : LTA	12,60
$x_5$ : ROAA	0,55%	$x_5$ : ROAA	-1,69%
$x_6$ : ROAE	6,18%	$x_6$ : ROAE	-14,05%
$x_7$ : IE II	40,21%	$x_7$ : IE II	49,08%
$x_9$ : PE OI	22,93%	$x_9$ : PE OI	24,02%
$x_{10}$ : PL GL	3,30%	$x_{10}$ : PL GL	10,57%
$x_{14}$ : EQ TA	9,32%	$x_{14}$ : EQ TA	9,48%

V tabulce 4.14 jsou uvedeny střední hodnoty pravděpodobností úpadků pro obě skupiny bank v kontrolním vzorku, a to pro všechny tři odhadnuté modely.

Tab. 4.14 Střední hodnoty PD bank z kontrolního vzorku pro jednotlivé modely

	Non - default banky	Default banky
	<i>E (PD)</i>	<i>E (PD)</i>
<b>Logit model</b>	14,47%	90,54%
<b>Probit model</b>	26,37%	75,38%
<b>LDA model</b>	7,39%	34,64%

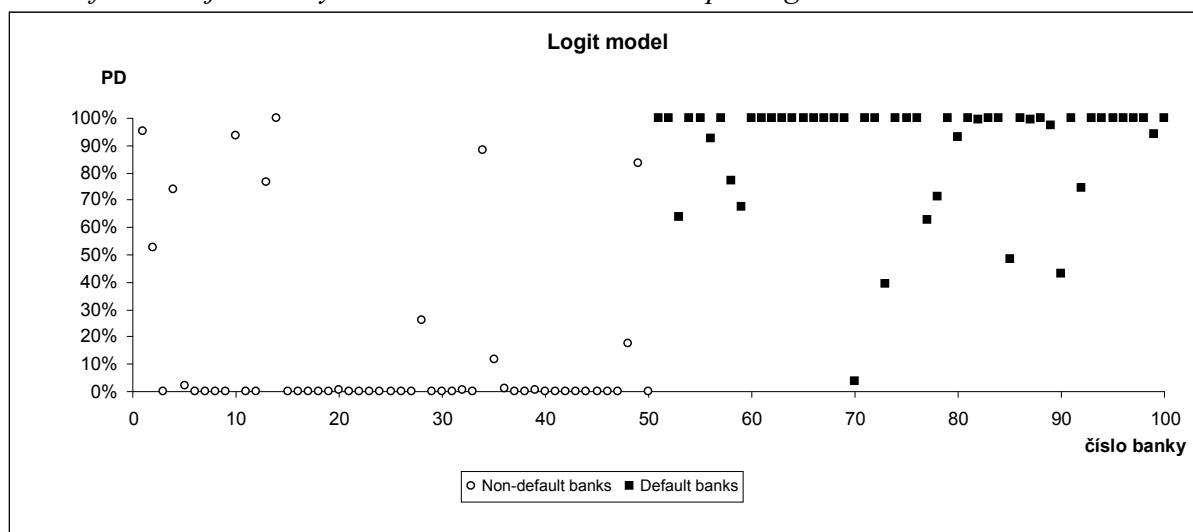
Z této tabulky je patrné, že dle aplikace odhadnutých modelů na kontrolní vzorek bank (banky v tomto vzorku byly vybrány náhodně na základě veřejně dostupných informací) se jeví jako nejlepší model pro predikci úpadku **logit model**. Oproti probit modelu tento model předpovídá úpadek bank lépe jak pro *non-default banky* (střední hodnota pravděpodobnosti úpadku je 14,47% oproti 26,37% u probit modelu) tak pro *default banky* (90,54% oproti 75,38% u probit modelu). Logit model tedy na tomto kontrolním vzorku bank vykazuje lepší výsledky než probit model navzdory tomu, že obsahuje o jeden ukazatel méně (konkrétně ukazatele LTA, ROAE a PL GL oproti ukazatelům LTA, ROAE, PE OI a EQ TA u probit modelu).

Model sestavený na základě lineární diskriminační analýzy předpovídá úpadek nejlépe ze všech tří odhadnutých modelů pro *non-default banky* (střední hodnota pravděpodobnosti úpadku činí pouhých 7,39%). Avšak tento model se při aplikaci na kontrolním vzorku bank vůbec neosvědčil při predikování úpadku u *default bank* (střední hodnota pravděpodobnosti úpadku je u tohoto modelu 34,64% oproti 90,54% u logit modelu a 75,38% u probit modelu). Tato skutečnost může být způsobena jeho nižší vypovídací schopností (viz 4.5). Z tohoto důvodu můžeme říct, že tento model je ze všech tří odhadnutých modelů nejméně vhodný, neboť není schopen spolehlivě předpovídat úpadek bank.

Tyto závěry jsou podloženy rovněž grafickým znázorněním na následujících třech grafech (Grafy 4.6 až 4.8), kde jsou zobrazeny pravděpodobnosti úpadku jednotlivých bank v kontrolním vzorku. Na ose x je znovu pořadové číslo banky a na ose y pravděpodobnost úpadku.

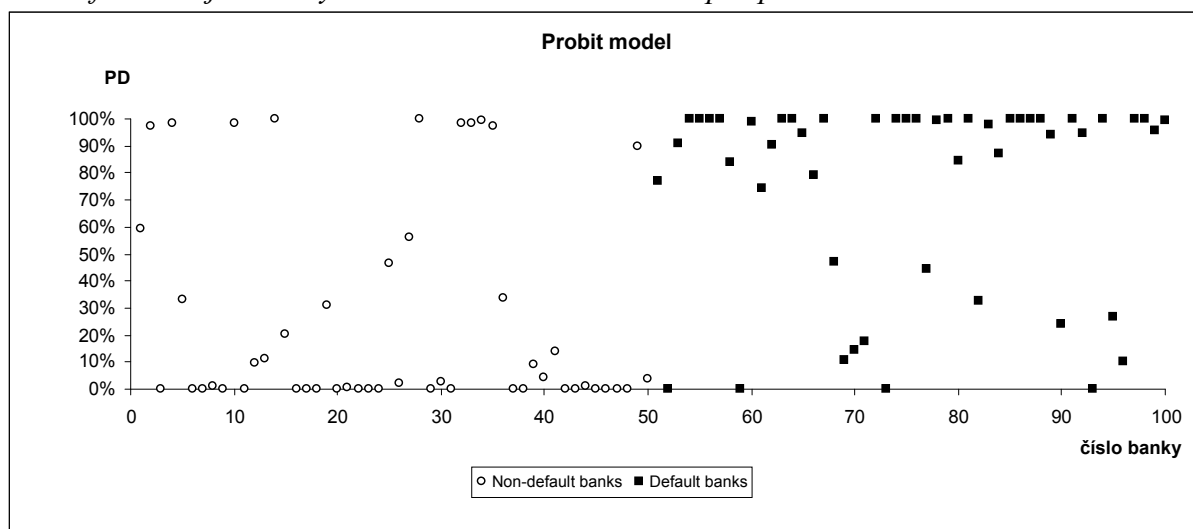


Graf 4.6 PD jednotlivých bank v kontrolním vzorku pro logit model



Z tohoto grafu lze vyčíst, že logit model označil správně<sup>2</sup> jak většinu non-default bank ze skupiny AA (pravděpodobnost úpadku nad 50% má v této skupině 8 bank), tak i default bank ze skupiny BB (pravděpodobnost úpadku pod 50% mají pouze 4 banky).

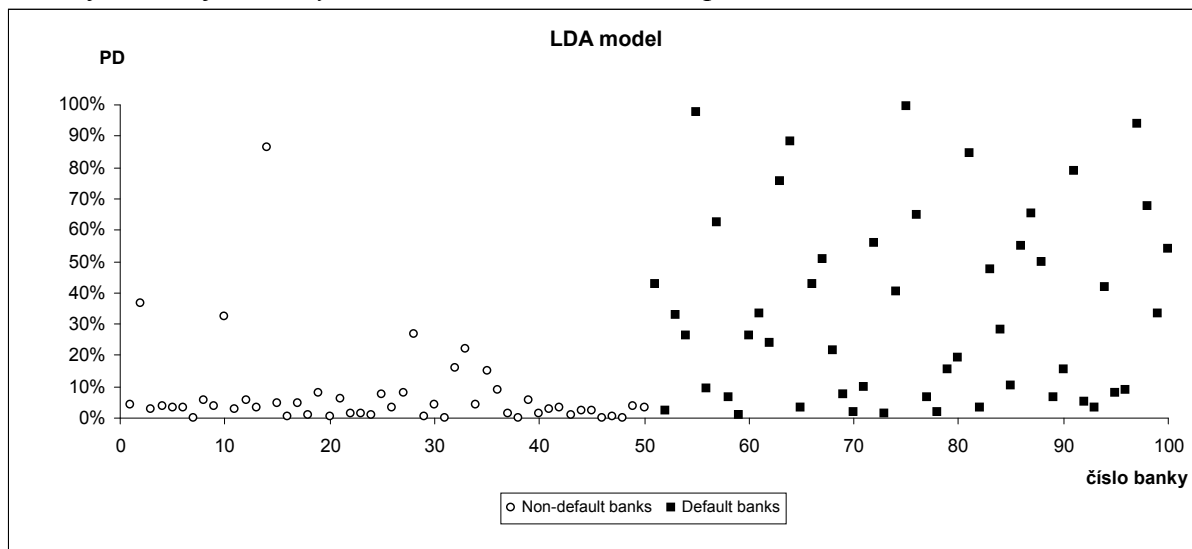
Graf 4.7 PD jednotlivých bank v kontrolním vzorku pro probit model



V grafu 4.7 vidíme, že v případě probit modelu se nad 50% hranicí pravděpodobnosti úpadku pohybuje již 10 bank ze skupiny AA a pod touto hranicí 13 bank ze skupiny BB.

<sup>2</sup> V případě, že si zvolíme 50% hranici pravděpodobnosti úpadku jako kritérium, kdy hodnoty pod touto hranicí pro non-default banky ve skupině AA a hodnoty nad touto hranicí pro default banky ve skupině BB znamenají správné zařazení bank do daných skupin.

Graf 4.8 PD jednotlivých bank v kontrolním vzorku pro LDA model



Z poslední grafu, na kterém jsou znázorněny výsledky pro LDA model, je patrné, že pravděpodobnost úpadku nad 50% má pouze jediná banka ze skupiny AA, ovšem ze skupiny BB se pod touto hranicí vyskytuje již 34 bank, což jednoznačně potvrzuje již dříve zmíněný závěr, že tento model není schopen spolehlivě predikovat úpadek bank.

## 5 Závěr

Bez správného určení pravděpodobnosti úpadku potenciálního dlužníka se neobejde žádná z významných oblastí řízení kreditního rizika. Při vzniku nedávné globální ekonomické krize hrálo zásadní roli zejména špatné ocenění kreditních derivátů, což mělo za následek kolaps finančního systému. Z těchto důvodů tato práce byla věnována odhadu pravděpodobnosti úpadku jakožto klíčového parametru v oblasti řízení kreditních rizik.

Cílem práce pak bylo odhadnout, verifikovat a srovnat modely pro určování pravděpodobnosti úpadku amerických komerčních bank a tyto následně aplikovat na kontrolní vzorek s cílem určit nejvhodnější z odhadnutých modelů pro predikci úpadku.

Ve druhé kapitole byly charakterizovány kreditní riziko a základní typy modelů pro predikci selhání. Na konci této kapitoly byl uveden výčet a stručný popis dnešních komerčních modelů. Součástí teoretické části práce byla i třetí kapitola. Ta byla zaměřena na detailní popis vybraných typů skóringových modelů. V této kapitole byla také uvedena charakteristika ukazatelů použitých při sestavování modelů v aplikační části.

V aplikační části práce (čtvrtá kapitola) byla představená teoreticko-metodologická východiska aplikována na vzorek bank. Na začátku této kapitoly byly charakterizovány vzorek bank a finanční ukazatele, jež posloužily jako vstupní data. Poté následoval samotný odhad jednotlivých modelů včetně statistické verifikace parametrů. Celkem byly odhadnuty tři predikční modely: logit model, probit model a LDA model. Všechny tyto modely patří do skupiny skóringových modelů. Logit a probit model se řadí k regresním modelům a liší se tím, že zatímco se v případě logit modelu používá logistická transformace, u probit modelu se aplikuje kumulativní distribuční funkce normovaného normálního rozdělení. LDA model se pak řadí do skupiny diskriminační analýzy. Součástí této kapitoly také bylo grafické znázornění pravděpodobností úpadků jednotlivých bank ze vzorku pro všechny tři odhadnuté modely. Dále bylo provedeno srovnání těchto modelů a jejich následná aplikace na kontrolní vzorek bank s cílem určit nejlepší model pro predikci úpadku.

Sestavené modely byly srovnány v podkapitole (4.5). Kromě základních rovnic jednotlivých modelů zde byly uvedeny počet a seznam jednotlivých finančních ukazatelů obsažených v příslušných modelech, vypovídací schopnost jednotlivých modelů a střední

hodnoty pravděpodobnosti úpadku pro obě skupiny bank. Logit a probit modely se ukázaly jako velice podobné jak ve vypovídací schopnosti (96,30% a 95,85%), tak ve středních hodnotách pravděpodobnosti úpadku (grafické srovnání závislosti mezi skórem a pravděpodobností úpadku u těchto dvou modelů bylo taktéž součástí této části práce). Probit model však obsahuje o jeden ukazatel navíc. Oproti těmto dvěma modelům obsahuje LDA model již pět ukazatelů, ale i přesto je jeho vypovídací schopnost nejnižší (78,44%).

V závěrečné podkapitole (4.6) byla provedena již zmíněná aplikace odhadnutých modelů na kontrolním vzorku bank, ze které vyplývá, že jako nejlepší model pro predikci úpadku se z těchto tří modelů jeví **logit model**. Tento model obsahuje 3 finanční ukazatele: logaritmus celkových aktiv (*LTA*), rentabilitu průměrného vlastního kapitálu (*ROAE*) a ukazatel poměřující problémové půjčky k celkovým půjčkám (*PL GL*). V případě tohoto modelu činí střední hodnota pravděpodobnosti úpadku pro non-default banky z kontrolního vzorku 14,47% a pro default banky 90,54%.

## Seznam použité literatury

### Knihy:

- [1] BALTHAZAR, L. *From Basel 1 to Basel 3: The Integration of State-of-the-Art Risk Modeling in Banking Regulation*. 1st ed. New York: Palgrave Macmillan, 2006. 294 s. ISBN 978-1-4039-4888-5.
- [2] BLUHM, CH.; OVERBECK, L.; WAGNER, CH. *An Introduction to Credit Risk Modeling*. 1st ed. Boca Raton: CRC Press LLC, 2003. 286 s. ISBN 1-58488-326-X.
- [3] DUFFIE, D.; SINGLETON, K. J. *Credit Risk: Pricing, Measurement, and Management*. 1st ed. New Jersey: Princeton University Press, 2003. 396 s. ISBN 0-691-09046-7.
- [4] GURNÝ, M. *Analýza finančního zdraví podniků pomocí matematicko-statistických metod*. Bakalářská práce, VŠB-TU Ostrava, ekonomická fakulta, katedra financí, 2009.
- [5] HOSMER, D. W.; LEMESHOW, S. *Applied Logistic Regression*. 2nd ed. New York: Wiley, 2000. 375 s. ISBN 978-0-471-35632-8.
- [6] RESTI, A.; SIRONI, A. *Risk management and Shareholders' value in banking*. 1st ed. Chichester: Wiley, 2007. 782 s. ISBN 978-0-470-02978-7.
- [7] *STATA Base Reference Manual (release 11)*. College Station: Stata Press, 2009. 2117 s. ISBN 1-59718-066-1.
- [8] ZMEŠKAL, Z. a kol. *Finanční modely*. 2. vyd. Praha: Ekopress, 2004. 236 s. ISBN 80-86119-87-4.

### Články:

GURNÝ, P., GURNÝ, M. Comparison of the credit scoring models on PD estimation of US banks, *Mathematical Method in Economics*, 2010, s. 188-194. ISBN 978-80-7394-218-2.

GURNÝ, P., GURNÝ, M. Logit and probit model within estimation of US banks PD, *Proceedings of the 47th Meeting of the Euro working group of financial modelling*, 2010, s. 73-80. ISBN 978-80-248-2351-5.

KARMINSKY, A., PERESETSKY, A.: Models for Moody's bank rating, *BOFIT Discussion Papers 17*, 2007.

### www stránky:

<http://www.federalreserve.gov/releases/lbr/current/default.htm>

<http://www.fdic.gov/bank/individual/failed/banklist.html>

<http://www.failedbankreporter.com/>

[http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/Psuedo\\_RSquareds.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm)

[http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/nested\\_tests.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/nested_tests.htm)

## Seznam zkratek a symbolů

$A$	skupina „zdravých“ subjektů
$A_T$	hodnota aktiv v čase $T$
AIBL	průměrná úročená pasiva ( <i>interest bearing liabilities</i> )
AIEA	průměrná úročená aktiva ( <i>average interest earnings assets</i> )
$B$	standardní Brownův pohyb; skupina subjektů v úpadku
CAR	celkový kapitál / rizikově vážená aktiva
CIBL	nákladové úroky / průměrná úročená pasiva
CIR	<i>cost to income ratio</i>
$D$	celkové vklady ( <i>total deposits</i> ), nominální hodnota dluhopisu
EAD	očekávaná hodnota expozice v případě selhání ( <i>exposure at default</i> )
EDF	očekávané frekvence úpadku ( <i>expected default frequencies</i> )
EL	očekávaní ztráta ( <i>expected loss</i> )
EQ	vlastní kapitál ( <i>shareholder's equity</i> )
ESS	rozptyl vysvětlený modelem ( <i>explained sum of squares</i> )
$E(Y x)$	podmíněná střední hodnota
$F$	$F$ -statistika
$FISH$	distribuční funkce Fisherova rozdělení
$FISH_{df_{ESS}; df_{RSS}}^{-1}$	inverzní funkce na hladině pravděpodobnosti $\alpha$
$F^{krit}$	$F$ -statistika kritická
$F^{vyp}$	$F$ -statistika vypočtená
GL	celkové půjčky ( <i>gross loans</i> )
$H_0$	nulová hypotéza
$H_A$	alternativní hypotéza
IE	nákladové úroky ( <i>interest expenses</i> )
II	výnosové úroky ( <i>interest income</i> )
$L_0$	log-likelihood více restriktivního modelu (modelu obsahujícího pouze konstantu)
$L_1$	log-likelihood méně restriktivního modelu („plného“ modelu)
LDA	lineární diskriminační analýza
LGD	očekávaná ztrátovost v případě selhání ( <i>loss given default</i> )
LLR	rezervy na ztrátové půjčky ( <i>loan loss reserve</i> )
LR	statistika likelihood-ratio testu

LTA	logaritmus celkových aktiv ( <i>logarithm of total assets</i> )
$L(\beta)$	logaritmus věrohodnostní funkce
MS	průměrný rozptyl ( <i>mean square</i> )
NIM	čistá úroková marže
$N(x)$	pravděpodobnost, že proměnná z normovaného normálního rozdělení je menší než $x$
OE	provozní náklady ( <i>operating expenses</i> )
OI	provozní výnosy ( <i>operating income</i> )
OLS	metoda nejmenších čtverců ( <i>ordinary least squares method</i> )
OTC	<i>over the counter</i>
PD	pravděpodobnost úpadku ( <i>probability of default</i> )
PE	osobní náklady ( <i>personnel expenses</i> )
PL	problémové půjčky ( <i>problem loans</i> )
QDA	kvadratická diskriminační analýza
$R^2$	koeficient determinace
ROAA	rentabilita průměrných aktiv ( <i>return on average assets</i> )
ROAE	rentabilita průměrného kapitálu ( <i>return on average equity</i> )
RR	míra návratnosti ( <i>recovery rate</i> )
RSS	rozptyl přiřazený reziduálnímu rozptylu nevysvětlenému modelem ( <i>residual sum of squares</i> )
RWA	rizikově vážená aktiva ( <i>risk weighted assets</i> )
SE	směrodatná odchylka ( <i>standard error</i> )
$ST$	distribuční funkce Studentova rozdělení
$ST_{\alpha/2;df}^{-1}$	inverzní funkce na hladině pravděpodobnosti $\alpha/2$ a stupňů volnosti $df$
$T$	doba splatnosti
T1	tier 1 ratio
T1 C	kapitál tier 1 ( <i>tier 1, core capital</i> )
TA	celková aktiva ( <i>total assets</i> )
TC	celkový kapitál ( <i>total capital</i> )
$W$	statistika Waldova testu
$X$	Brownův pohyb s rozptylem rovným jedné konstantním přírůstkem $m$
YAEA	výnosové úroky / průměrná úročená aktiva
$g(x)$	nezávisle proměnná



$m$	rychlost změny průměrné vzdálenosti k úpadku
$n$	počet nezávislých pozorování; počet subjektů
$n_A$	počet subjektů ve skupině A
$n_B$	počet subjektů ve skupině B
$p$	počet nezávislých proměnných
$p_E$	hladina významnosti pro přidání proměnné do modelu
$p_R$	hladina významnosti pro vyloučení proměnné z modelu
$p(t)$	pravděpodobnost, že dlužník neselže po $t$ let
$t_{df}$	$t$ -statistika s $df$ stupni volnosti
$t^{\text{krit}}$	$t$ -statistika kritická
$t^{\text{vyp}}$	$t$ -statistika vypočtená
$x$	nezávisle proměnná
$x'$	vektor $p$ nezávislých proměnných
$x_A$	vektor středních hodnot pro skupinu A
$x_B$	vektor středních hodnot pro skupinu B
$y$	závisle proměnná; podmíněná střední hodnota v případě probit regrese
$z$	střední hodnota z skóre pro celý vzorek subjektů
$z_{1-\alpha/2}$	horní kritická hodnota z normovaného normálního rozdělení na hladině významnosti $\alpha/2$
$z_A$	střední hodnota z skóre pro skupinu A
$z_B$	střední hodnota z skóre pro skupinu B
$z_i$	z skóre pro $i$ -tý subjekt
$\Lambda$	Wilksovo Lambda
$\Sigma$	„průměrná“ kovarianční matice
$\Sigma_A$	kovarianční matice skupiny A
$\Sigma_B$	kovarianční matice skupiny B
$\alpha$	hladina významnosti
$\beta_0$	úrovňová konstanta
$\beta_j$	parametr pro $j$ -tou nezávisle proměnnou
$\beta'$	vektor parametrů

$\hat{\beta}$	odhadnutý parametr (odhady maximální věrohodnosti)
$\gamma$	dividendový výnos; koeficient gama; vektor koeficientu gama
$\gamma'$	upravený vektor koeficientů gama
$\varepsilon$	reziduální odchylka
$\lambda$	konstantní střední hodnota míry výskytu selhání (intenzita)
$\mu$	průměrná míra rentability aktiv
$\pi_B$	pravděpodobnost selhání nezávislá na konkrétní charakteristice subjektu
$\pi(x)$	podmíněná střední hodnota v případě logistické regrese
$\hat{\pi}(x)$	predikované hodnoty logit modelu
$\sigma$	volatilita aktiv
$\tau$	čas prvního výskytu selhání
$\chi^2$	chi-kvadrát rozdělení

## Seznam obrázků, tabulek a grafů

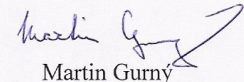
Obr. 2.1	Zjednodušené schéma struktury neuronové sítě
Obr. 2.2	Schéma genetického algoritmu
Tab. 2.1	Přehled skóringových modelů
Tab. 2.2	Přehled dnešních komerčních modelů
Tab. 4.1	Přehled použitých ukazatelů a jejich zařazení do základních oblastí
Tab. 4.2	Střední hodnota finančních ukazatelů pro non-default a default banky
Tab. 4.3	Odhadnutý logit model
Tab. 4.4	Waldův test logit modelu
Tab. 4.5	Odhadnuté střední hodnoty PD pro logit model
Tab. 4.6	Odhadnutý probit model
Tab. 4.7	Waldův test probit modelu
Tab. 4.8	Odhadnuté střední hodnoty PD pro probit model
Tab. 4.9	t-test LDA modelu
Tab. 4.10	F-test LDA modelu
Tab. 4.11	Odhadnuté střední hodnoty z skóre a PD pro LDA model
Tab. 4.12	Srovnání jednotlivých odhadnutých modelů
Tab. 4.13	Střední hodnoty vybraných fin. ukazatelů kontrolního vzorku bank
Tab. 4.14	Střední hodnoty PD bank z kontrolního vzorku pro jednotlivé modely
Graf 4.1	Odhadnuté PD jednotlivých bank logit modelu
Graf 4.2	Odhadnuté PD jednotlivých bank probit modelu
Graf 4.3	Odhadnuté PD jednotlivých bank LDA modelu
Graf 4.4	Srovnání logit a probit modelu
Graf 4.5	Detail funkcí logit a probit modelu
Graf 4.6	PD jednotlivých bank v kontrolním vzorku pro logit model
Graf 4.7	PD jednotlivých bank v kontrolním vzorku pro probit model
Graf 4.8	PD jednotlivých bank v kontrolním vzorku pro LDA model

## Prohlášení o využití výsledků diplomové práce

Prohlašuji, že

- jsem byl seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. – autorský zákon, zejména § 35 – užití díla v rámci občanských a náboženských obřadů, v rámci školních představení a užití díla školního a § 60 – školní dílo;
- beru na vědomí, že Vysoká škola báňská – Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užít (§ 35 odst. 3);
- souhlasím s tím, že diplomová práce bude v elektronické podobě archivována v Ústřední knihovně VŠB-TUO a jeden výtisk bude uložen u vedoucího diplomové práce. Souhlasím s tím, že bibliografické údaje o diplomové práci budou zveřejněny v informačním systému VŠB-TUO;
- bylo sjednáno, že s VŠB-TUO, v případě zájmu z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona;
- bylo sjednáno, že užít své dílo, diplomovou práci, nebo poskytnout licenci k jejímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Ostravě dne 29. dubna 2011



Martin Gurný  
jméno a příjmení studenta

Adresa trvalého pobytu studenta:  
Mojmírovců 1248/38, Ostrava 9, 709 00.

## Seznam příloh

Příloha č. 1	Seznam bank v původním vzorku
Příloha č. 2	Vypočtené hodnoty 16 finančních ukazatelů pro jednotlivé banky v původním vzorku
Příloha č. 3	Korelační matice použitých finančních ukazatelů
Příloha č. 4	Seznam bank v kontrolním vzorku
Příloha č. 5	Vypočtené hodnoty vybraných finančních ukazatelů pro jednotlivé banky v kontrolním vzorku